



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 105282

TO: Vanessa L Ford
Location: CM-1/8A16/8E12
Art Unit: 1645
Monday, October 06, 2003

Case Serial Number: 09765739

From: Alex Waclawiw
Location: Biotech-Chem Library
CM1-6A02
Phone: 308-4491

Alexandra.waclawiw@uspto.gov

Search Notes

1

From: Chan, Christina
Sent: Friday, October 03, 2003 10:25 AM
To: Ford, Vanessa; STIC-Biotech/ChemLib
Subject: RE: IN RE: 09/765, 739 Sequence search

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

-----Original Message-----

From: Ford, Vanessa
Sent: Thursday, October 02, 2003 6:38 PM
To: Chan, Christina
Subject: IN RE: 09/765, 739 Sequence search

Search SEQ ID NOs: 1-7

Please include interference searches. Please rush.

Vanessa L. Ford
Biotechnology Patent Examiner
Office: CM1 8A16
Mailbox: CM1 8E12
Phone: 703.308.4735
Art Unit: 1645

Point of Contact:
Alexandra Wacławiw
Technical Info. Specialist
CM1 6A02 Tel: 308-4491

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10-6-03
Date Completed: 10-6-03
Searcher Prep/Review: 9
Clerical: 18
Online time: 35

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 7
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): CompuLink

RECEIVED
OCT - 3 2003
STIC-Biotech/ChemLib



2.1

2.1.1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 24.4478 seconds
(without alignments)
189.995 Million cell updates/sec

Title: US-09-765-739A-6

Perfect score: 97

Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_23:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	278	2	O52106 ehrlichia c
2	97	100.0	278	2	Q8G8Q5 ehrlichia c
3	88	90.7	280	2	O52104 ehrlichia c
4	88	90.7	280	2	Q8G8Q5 ehrlichia c
5	88	90.7	280	2	Q8G8J3 ehrlichia c
6	76	78.4	276	2	Q8G948 ehrlichia c
7	76	78.4	277	2	Q8G8W7 ehrlichia c
8	74	76.3	277	2	Q8G8U1 ehrlichia c
9	59	60.8	246	2	Q8R335 ehrlichia c
10	59	60.8	275	2	Q33DD4 ehrlichia c
11	59	60.8	276	2	Q33DD1 ehrlichia c
12	59	60.8	276	2	Q85817 ehrlichia c
13	59	60.8	276	2	Q8G8U0 ehrlichia c
14	59	60.8	280	2	O52107 ehrlichia c
15	59	60.8	288	2	Q92GJ2 ehrlichia c
16	58	59.8	281	2	Q93DD2 ehrlichia c

90 43 44.3 279 2 Q8G8Q1 Q898q1 ehrlichia c
 91 43 44.3 298 16 P73094 P73094 synchocyst
 92 43 44.3 389 12 Q69554 Q69554 human herpe
 93 43 44.3 389 12 Q9WT45 Q9WT45 human herpe
 94 43 44.3 407 16 Q8EG37 Q8EG37 shewanella
 95 43 44.3 412 12 Q69043 Q69043 human herpe
 96 43 44.3 853 15 Q8UPN0 Q8UPN0 human immun
 97 42.5 43.8 370 3 Q14341 Q14341 schizosacch
 98 42 43.3 91 2 Q915F8 Q915F8 salmonella
 99 42 43.3 92 16 Q935M7 Q935M7 salmonella
 100 42 43.3 190 16 Q97RX1 Q97RX1 streptococc

ALIGNMENTS

RESULT 1

OS2106 PRELIMINARY; PRT; 278 AA.
 AC 052106; 085357; PubMed=9423849;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OMP-1E (28kDa outer membrane protein gene 17).
 GN OMP-1E.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=98084465; PubMed=9423849;
 RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 RT are encoded by a polymorphic multigene family.";
 RL Infect. Immun. 66:132-139(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=98321180; PubMed=9647746;
 RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,
 RA Alleman A.R.;
 RT "Molecular characterization of a 28 kDa surface antigen gene family of
 RT the tribe Ehrlichiae.";
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oscitola;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 RT and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; U72291; AAC02919.1; -;
 DR EMBL; AF479833; AAC02919.1; -;
 DR EMBL; AF479834; AAC02936.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1. E321E3CA259B87FD CRC64;
 SQ SEQUENCE 278 AA; 30543 MW; E321E3CA259B87FD CRC64;

Query Match 100.0%; Score 97; DB 2; Length 278;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
 |||||

DB 60 NPTVALYGLKQDWEGISS 77
 |||||

RESULT 2

Q8G8Q5

ID Q8G8Q5 PRELIMINARY; PRT; 278 AA.
 AC 08G8Q5;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE 28kDa outer membrane protein gene 17.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Liberty, and Jax;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 RT and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479839; AAC02963.1; -;
 DR EMBL; AF479840; AAC02969.1; -;
 SQ SEQUENCE 278 AA; 30381 MW; 7626778E5798D6B7 CRC64;
 Query Match 100.0%; Score 97; DB 2; Length 278;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWEGISS 18
 |||||

ID Q8G8Q5 PRELIMINARY; PRT; 280 AA.
 AC 052104; 085357; PubMed=9423849;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OMP-1C (28kDa outer membrane protein 15).
 GN OMP-1C.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=98084465; PubMed=9423849;
 RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 RT are encoded by a polymorphic multigene family.";
 RL Infect. Immun. 66:132-139(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;
 RX MEDLINE=98321180; PubMed=9647746;
 RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,
 RA Alleman A.R.;
 RT "Molecular characterization of a 28 kDa surface antigen gene family of
 RT the tribe Ehrlichiae.";
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;
 RX MEDLINE=22384137; PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 RT and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; U72291; AAC02937.1; -;
 DR EMBL; AF479833; AAC02930.1; -;

```
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30322 MW; BD835E491086DF01 CRC64;

Query Match          90.7%; Score 88; DB 2; Length 280;
Best Local Similarity 83.3%; Pred. No. 5.3e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NPTVALYGLKQDWNGVSA 77

RESULT 4
QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NPTVALYGLKQDWNGVSA 77

ID Q8GGU5 PRELIMINARY; PRT; 280 AA.
AC Q8GGU5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 15.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479834; AAO12934.1; -.
SQ SEQUENCE 280 AA; 30332 MW; BD835D792386DF01 CRC64;

Query Match          90.7%; Score 88; DB 2; Length 280;
Best Local Similarity 83.3%; Pred. No. 5.3e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NPTVALYGLKQDWNGVSA 77

RESULT 5
QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NPTVALYGLKQDWNGVSA 77

ID Q8GJ3 PRELIMINARY; PRT; 280 AA.
AC Q8GJ3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 15.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479839; AAO12961.1; -.
SQ SEQUENCE 280 AA; 30721 MW; A36BE26DACA2C0D CRC64;

Query Match          90.7%; Score 88; DB 2; Length 280;
Best Local Similarity 83.3%; Pred. No. 5.3e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NPTVALYGLKQDWNGVSA 77

RESULT 6
QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NPTVALYGLKQDWNGVSA 77

ID Q8G948 PRELIMINARY; PRT; 276 AA.
AC Q8G948;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 15.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=West Paces, Heartland, St. Vincent, and Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479835; AAO12940.1; -.
DR EMBL; AF479836; AAO12945.1; -.
DR EMBL; AF479837; AAO12950.1; -.
DR EMBL; AF479838; AAO12955.1; -.
SQ SEQUENCE 276 AA; 30316 MW; OD6F5353F9C0F17C CRC64;

Query Match          78.4%; Score 76; DB 2; Length 276;
Best Local Similarity 77.8%; Pred. No. 0.00044;
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NATVALYGLKQDWNGASA 77

RESULT 7
QY 1 NPTVALYGLKQDWEGISS 18
DB 60 NATVALYGLKQDWNGASA 77

ID Q8G8W7 PRELIMINARY; PRT; 277 AA.
AC Q8G8W7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479835; AAO12942.1; -.
DR EMBL; AF479836; AAO12947.1; -.
DR EMBL; AF479837; AAO12952.1; -.
SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

Query Match          78.4%; Score 76; DB 2; Length 277;
Best Local Similarity 77.8%; Pred. No. 0.00044;
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 NPTVALYGLKQDWEGISS 18
ID 1 NPTVALYGLKQDWEGISS 18
Db 60 NATVALYGLKQDWNGASA 77

RESULT 8
Q8GGU1 PRELIMINARY; PRT; 277 AA.
AC Q8GGU1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479838; AAO12957.1; -.
SQ SEQUENCE 277 AA; 30294 MW; 0CE7EDB51F2D854E CRC64;

Query Match 76.3%; Score 74; DB 2; Length 277;
Best Local Similarity 72.2%; Pred. No. 0.0093;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
ID 1 NPTVALYGLKQDWEGISS 18
Db 60 NATVALYGLKQDWNGASA 77

RESULT 9
Q9RH35 PRELIMINARY; PRT; 246 AA.
AC Q9RH35;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (Fragment).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St.Vincent;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL; AF077735; AAC31548.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 60.8%; Score 59; DB 2; Length 246;
Best Local Similarity 60.0%; Pred. No. 0.21;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
ID 1 NPTVALYGLKQDWEG 15
Db 29 NTTAGVFLGKQDWG 43

RESULT 10
Q93DD4 PRELIMINARY; PRT; 275 AA.
AC Q93DD4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V2;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393389; AAL12919.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 29974 MW; 2ECF2F988B2E9D9 CRC64;

Query Match 60.8%; Score 59; DB 2; Length 275;
Best Local Similarity 60.0%; Pred. No. 0.24;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
ID 1 NPTVALYGLKQDWEG 15
Db 59 NTTAGVFLGKQDWG 73

RESULT 11
Q93DD1 PRELIMINARY; PRT; 276 AA.
AC Q93DD1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393393; AAL12923.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30028 MW; 2D7143AFCBFF2EBE CRC64;

Query Match 60.8%; Score 59; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.24;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
ID 1 NPTVALYGLKQDWEG 15
Db 59 NTTAGVFLGKQDWG 73

RESULT 12
Q85817 PRELIMINARY; PRT; 276 AA.
ID Q85817;
AC Q85817;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
```

RL	Infect. Immun.	71:187-195(2003).
RQ	EMBL; AF479838; AAO12958.1;	-
SQ	SEQUENCE	276 AA; 30045 MW; 8E3C9719D3C67A64 CRC64;
Query Match	60.8%;	Score 59; DB 2; Length 276;
Best Local Similarity	60.0%;	Pred.No. 0.24;
Matches	9; Conservative	3; Mismatches 3; Indels 0; Gaps 0;
QY	1 NPTVALYGLKQDWEG 15	
DG	I I : I	
DB	59 NTTAGVGLKQWDG 73	
RESULT 14		
O52107	PRELIMINARY;	PRT; 280 AA.
ID	O52107	
AC	O52107	
DT	01-JUN-1998 (TEMBLrel. 06, Created)	
DT	01-JUN-1998 (TEMBLrel. 06, Last sequence update)	
DT	01-MAR-2003 (TEMBLrel. 23, Last annotation update)	
DE	Major outer membrane protein OMP-1F (28kDa outer membrane protein gene	
DE	18).	
GN	OMP-1F.	
OS	Ehrlichia chaffeensis.	
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;	
OC	Anaplasmataceae; Ehrlichia.	
OX	NCBI_TaxID=945;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-Arkansas;	
RC	MEDLINE=98084465; PubMed=9423849;	
RA	Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;	
RT	"Immunodominant major outer membrane proteins of Ehrlichia chaffeensis	
RT	are encoded by a polymorphic multigene family.";	
RT	Infect. Immun. 66:132-139(1998).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-Arkansas;	
RC	MEDLINE=21153566; PubMed=11254561;	
RA	Ohashi N., Rikihisa Y., Unver A.;	
RT	"Analysis of transcriptionally Active Gene Clusters of Major Outer	
RT	Membrane Protein Multigene Family in Ehrlichia canis and E.	
RT	chaffeensis.";	
RT	Infect. Immun. 69:2083-2091(2001).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-Arkansas;	
RC	MEDLINE=98321180; PubMed=9647746;	
RA	Reddy G.R., Sulsano C.R., Barbet A.F., Mahan S.M., Burridge M.J.,	
RA	Allaman A.R.;	
RT	"Molecular characterization of a 28 kDa surface antigen gene family of	
RT	the tribe Ehrlichiae.";	
RT	Biochem. Biophys. Res. Commun. 247:636-643(1998).	
RN	[4]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-Arkansas, and Osciola;	
RC	PubMed=12496165;	
RA	Cheng C., Paddock C.D., Ganta R.R.;	
RT	"Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined	
RT	by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes	
RT	and Other Regions of the Genome.";	
RT	Infect. Immun. 71:187-195(2003).	
DR	EMBL; U72291; AAC02940.1;	
DR	EMBL; AF479833; AAO12931.1;	
DR	EMBL; AF479834; AAO12937.1;	
DR	InterPro: IPR002566; Surface_Ag_msp4.	
DR	Pfam: PF01617; Surface_Ag_2; 1.	
SQ	SEQUENCE	280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;
Query Match	60.8%;	Score 59; DB 2; Length 280;
Best Local Similarity	60.0%;	Pred.No. 0.24;
Matches	9; Conservative	3; Mismatches 3; Indels 0; Gaps 0;

```
OY 1 NPTVALYGLKQDWEG 15
Db 60 NTTGVFGLKQDWG 74

RESULT 15
Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
AC Q9ZGJ2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 30-kDa major outer membrane protein (P28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RA MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF078553; AAC68667.1; -.
DR EMBL; AF082744; AAG14362.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BF5E CRC64;

Query Match 60.8%; Score 59; DB 2; Length 288;
Best Local Similarity 60.0%; Pred. No. 0.25;
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 NPTVALYGLKQDWEG 15
Db 60 NTTGVFGLKQDWG 74

RESULT 16
Q93DD2 PRELIMINARY; PRT; 281 AA.
AC Q93DD2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lithonia, Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
```

RT the tribe Ehrlichiae.";
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
 DR EMBL: U72291; AAK28673.1; -
 DR EMBL: AF393388; AAL12918.1; -
 DR EMBL: AY117396; AAM77031.1; -
 DR EMBL: AF479833; AAO12932.1; -
 DR EMBL: AF479834; AAO12938.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 59.8%; Score 58; DB 2; Length 281;
 Best Local Similarity 60.0%; Pred. No. 0.35;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
 | | | : | | | : | | :
 DB 59 NTTGVFGLKQNDG 73

RESULT 18
 Q9AMF6 PRELIMINARY; PRT; 272 AA.
 ID Q9AMF6;
 AC Q9AMF6;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Major antigenic protein MAP1 (Fragment).
 GN MAP1
 OS Ehrlichia sp. 'South African canine'.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=152574;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Allsopp M.T.; Allsopp B.A.;
 RT "A novel Ehrlichia detected in dogs in South Africa.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF325176; AAK14320.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT NON_TER 1
 FT NON_TER 272
 SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

Query Match 57.7%; Score 56; DB 2; Length 272;
 Best Local Similarity 50.0%; Pred. No. 0.71;
 Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
 | | | : | | | : | | :
 DB 57 TKAVFGLKQDWGVKT 72

RESULT 19
 Q93E54 PRELIMINARY; PRT; 272 AA.
 ID Q93E54;
 AC Q93E54;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Major antigenic protein MAP1 (Fragment).
 OS Cowdria ruminantium.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=779;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pokoase;
 RX MEDLINE=21539003; PubMed=11682561;
 RA Allsopp M.T.E.P.; Dorfling C.M.; Maillard J.C.; Bensaid A.;
 RA Haydon D.T.; van Heerden H.; Allsopp B.A.;
 RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants

RT Are Not Geographically Constrained and Show No Evidence of Having
 RT Evolved under Positive Selection Pressure.";
 RL J. Clin. Microbiol. 39:4200-4203(2001).
 DR EMBL: AF368013; AAK98153.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT NON_TER 1
 FT NON_TER 272
 SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

Query Match 57.7%; Score 56; DB 2; Length 272;
 Best Local Similarity 50.0%; Pred. No. 0.71;
 Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
 | | | : | | | : | | :
 DB 57 TKAVFGLKQDWGVKT 72

RESULT 20
 Q9ADV3 PRELIMINARY; PRT; 280 AA.
 ID Q9ADV3;
 AC Q9ADV3;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
 DE Major outer membrane protein P30-2.
 GN P30-2.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Okiahoma;
 RX MEDLINE=98371112; PubMed=9705412;
 RA Ohashi N.; Unver A.; Zhi N.; Rikihisa Y.;
 RT "Cloning and characterization of multigenes encoding the
 immunodominant 30-kilodalton major outer membrane proteins of
 Ehrlichia canis and application of the recombinant protein for
 serodiagnosis.";
 RL J. Clin. Microbiol. 36:2671-2680(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Okiahoma;
 RX MEDLINE=21153566; PubMed=11254561;
 RA Ohashi N.; Rikihisa Y.; Unver A.;
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
 Membrane Protein Multigene Family in Ehrlichia canis and E.
 chaffeensis.";
 RL Infect. Immun. 69:2083-2091(2001).
 DR EMBL: AF078553; AAK28699.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match 57.7%; Score 56; DB 2; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.73;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
 | | | : | | | : | | :
 DB 59 NSTGVFGLKHDWNG 73

RESULT 21
 Q9F473 PRELIMINARY; PRT; 280 AA.
 ID Q9F473;
 AC Q9F473;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE P28-6.

GN 01-NOV-1996 (TRENBLrel. 01, Created)
 OS 01-DEC-2001 (TRENBLrel. 01, Last sequence update)
 OC Major antigenic protein.
 OX MAP1.
 RN [1]_TaxID=944;
 RP SEQUENCE FROM N.A.
 RC STRAIN-Jake;
 RX MEDLINE=99242757; PubMed=10225842;
 RA MCBride J.W., Yu X.J., Walker D.H.;
 RT "Molecular cloning of the gene for
 RT 28-kilodalton protein of Ehrlichia canis;
 RT antigen";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;
 Query Match
 Best Local Similarity 57.7%; Score 56; DB 2; Length 280;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NTVALYGLKQDWEIG 15
 DB 59 NSTVGLKQDWEIG 73
 RESULT 22
 ID Q9AFAL PRELIMINARY; PRT; 284 AA.
 AC Q9AFAL;
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DE Major antigenic protein 1.
 OS Cowdria ruminantium.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OX NCBI_TaxID=779;
 RP SEQUENCE FROM N.A.
 RC STRAIN-Ball-3;
 RX "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
 RT are not geographically constrained";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF355200; AAK27216.1;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 284 AA; 30720 MW; B0D3AEB9F9A809C1 CRC64;
 Query Match
 Best Local Similarity 57.7%; Score 56; DB 2; Length 284;
 Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWEIG 18
 DB 63 TRAVFGLKQDWEIG 78
 RESULT 23
 ID Q46327 PRELIMINARY; PRT; 284 AA.
 AC Q46327;
 DT 01-NOV-1996 (TRENBLrel. 01, Created)
 DT 01-DEC-2001 (TRENBLrel. 01, Last sequence update)
 DE Major antigenic protein.
 GN MAP1.
 OS Cowdria ruminantium.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OX NCBI_TaxID=779;
 RP SEQUENCE FROM N.A.
 RC STRAIN-Senegal stock;
 RX MEDLINE=94178956; PubMed=8132352;
 RA van Vliet A.H., Jongejan F., van Kleef M., van der Zeijst B.A.;
 RT "Molecular cloning, sequence analysis, and expression of the gene
 RT encoding the immunodominant 32-kilodalton protein of Cowdria
 RL ruminantium";
 RL Infect. Immun. 62:1451-1456(1994).
 DR EMBL: X74250; CAA52309.1;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 284 AA; 30634 MW; 53228A889D28BEB8 CRC64;
 Query Match
 Best Local Similarity 57.7%; Score 56; DB 2; Length 284;
 Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWEIG 18
 DB 63 TRAVFGLKQDWEIG 78
 RESULT 24
 ID Q93DD3 PRELIMINARY; PRT; 280 AA.
 AC Q93DD3;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DE Outer membrane protein p28.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OX NCBI_TaxID=945;
 RP SEQUENCE FROM N.A.
 RC STRAIN-V5;
 RX "Allele variation and patterns of transcription of the Ehrlichia
 RT chaffeensis p28 kDa outer membrane protein multigene family";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF393391; AAL12921.1;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30372 MW; C7BBC8710BC167E9 CRC64;
 Query Match
 Best Local Similarity 55.7%; Score 54; DB 2; Length 280;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWEIG 15
 DB 61 TAGVFLKQDWEIG 73
 RESULT 25
 ID Q9ZGM9 PRELIMINARY; PRT; 280 AA.
 AC Q9ZGM9

```
AC Q9ZGM9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077733; AAC31545.1; -
DR EMBL; AF077733; AAC31545.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Query Match 55.7%; Score 54; DB 2; Length 280;
Best Local Similarity 61.5%; Pred. No. 1.5;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
Db 61 TAGVFGKQDWGD 73

RESULT 26
O85816 PRELIMINARY; PRT; 280 AA.
AC O85816;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28kDa outer membrane protein gene
DE 19).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V8, and Va;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
```

```
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077732; AAC31545.1; -
DR EMBL; AF393394; AAL12924.1; -
DR EMBL; AF393390; AAL12920.1; -
DR EMBL; AF479839; AAO12964.1; -
DR EMBL; AF479840; AAO12970.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 55.7%; Score 54; DB 2; Length 280;
Best Local Similarity 61.5%; Pred. No. 1.5;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
Db 61 TAGVFGKQDWGD 73

RESULT 27
Q9F472 PRELIMINARY; PRT; 278 AA.
AC Q9F472;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-7.
GN P28-7.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF082744; AAC64550.2; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30455 MW; 2411CAAB4C56CA74 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
Db 62 TVGVFGKQDWGD 74

RESULT 28
Q9R8A8 PRELIMINARY; PRT; 278 AA.
AC Q9R8A8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
```

```

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OklaHoma;
RA MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082746; AAC64552.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
DB 62 TVGVFGLKHDWDG 74

RESULT 29
Q9R8A7 ID Q9R8A7 PRELIMINARY; PRT; 278 AA.
AC Q9R8A7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Demom;
RA MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082747; AAC64553.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
DB 62 TVGVFGLKHDWDG 74

RESULT 30
Q9R3J3 ID Q9R3J3 PRELIMINARY; PRT; 278 AA.
AC Q9R3J3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Demom;
RA MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082748; AAC64554.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
DB 62 TVGVFGLKHDWDG 74

RESULT 31
Q9R8A6 ID Q9R8A6 PRELIMINARY; PRT; 278 AA.
AC Q9R8A6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DJ;
RA MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082748; AAC64554.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
DB 62 TVGVFGLKHDWDG 74

RESULT 32
Q9R8A9 ID Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DJ;
RA MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082748; AAC64554.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
DB 62 TVGVFGLKHDWDG 74

```

OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Louisiana;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082745; AAC64551.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
Db 62 TVGVFGLKRDWDG 74

RESULT 33
Q9R8A5 PRELIMINARY; PRT; 278 AA.
AC Q9R8A5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Fuzzy;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082749; AAC64555.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 2.2;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
Db 62 TVGVFGLKRDWDG 74

RESULT 34
Q9ZGJ1 PRELIMINARY; PRT; 307 AA.
AC Q9ZGJ1;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Major outer membrane protein P30-1.
GN P30-1.

OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Oklahoma;
RX MEDLINE=98371112; PubMed=97054112;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Oklahoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAC68666.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 307 AA; 34108 MW; 9DA9FD63EBF8BC97 CRC64;

Query Match 54.6%; Score 53; DB 2; Length 307;
Best Local Similarity 61.5%; Pred. No. 2.4;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEG 15
Db 91 TVGVFGLKRDWDG 103

RESULT 35
Q9AF99 PRELIMINARY; PRT; 265 AA.
ID Q9AF99
AC Q9AF99;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE Major antigenic protein 1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Umpala;
RA Bensaïd A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
RT are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355202; AAK27218.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 265 265
SQ SEQUENCE 265 AA; 28388 MW; 1EA6DB25CCBF9634 CRC64;

Query Match 53.6%; Score 52; DB 2; Length 265;
Best Local Similarity 43.8%; Pred. No. 3;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
Db 45 TQTVFGLKRDWDGVKT 60

Search completed: October 6, 2003, 07:52:41

Mon Oct 6 09:43:10 2003

us-09-765-739a-6.rspt

Page 12

Job time : 26.4478 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:02 ; Search time 5.10448 Seconds
(without alignments)
165.831 Million cell updates/sec

Title: US-09-765-739A-6

Perfect score: 97

Sequence: 1 NPTVALXCLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	46	47.4	375	1	YBJF_SALTI	Q82841 salmonella
2	46	47.4	376	1	YBJF_SALTY	Q82qj5 salmonella
3	45	46.4	264	1	UPK2_RHILO	Q98nj1 rhizobium l
4	44	45.4	162	1	MEGL_VIBCH	Q9kpk1 vibrio chol
5	42	43.3	122	1	RT29_BOVIN	P82922 bos taurus
6	42	43.3	161	1	MENG_VERPE	Q82j77 yersinia pe
7	42	43.3	1023	1	SI23_PSEAM	P55019 pseudopleur
8	41.5	42.8	195	1	RU11_SPOFR	Q962u2 spodoptera
9	41	42.3	169	1	FLAV_ANASP	P11241 anabaena sp
10	41	42.3	368	1	ALR3_SALTI	Q82300 salmonella
11	41	42.3	452	1	YK97_MYCTU	Q10706 mycobacteri
12	41	42.3	475	1	MTHC_DROME	P83119 drosophila
13	41	42.3	544	1	PYRG_AZOB	P28595 azospirillum
14	41	42.3	741	1	FLO3_MOUSE	Q9r0el mus musculus
15	41	42.3	918	1	QAIS_NEUCR	P11637 neurospora
16	41	42.3	991	1	DHPI_SCHPO	P40848 schizosacch
17	40	41.2	169	1	FLAV_SYN7P	P10340 synechococc
18	40	41.2	196	1	NODA_MESPL	Q8vvh1 mesorhizobi
19	40	41.2	239	1	PYRF_BACSU	P25971 bacillus su
20	40	41.2	341	1	Y33B_MYCPN	P75302 mycoplasma
21	40	41.2	486	1	GLG2_RHIME	P56394 rhizobium m
22	40	41.2	661	1	PDAT_YEAST	Q40345 saccharomyc
23	40	41.2	734	1	PURL_ZYMO	Q9req6 zymomonas m
24	40	41.2	1060	1	NKCL_MANSE	Q25479 manduca sex
25	40	41.2	1220	1	DDX8_HUMAN	Q14562 homo sapien
26	40	41.2	4074	1	PKHD_HUMAN	Q8tc29 homo sapien
27	39.5	40.7	354	1	CARA_METJA	Q58425 methanococc
28	39	40.2	196	1	NODA_RHIS3	P72329 rhizobium s
29	39	40.2	213	1	RS3_OCEIH	P59182 oceanobacil
30	39	40.2	226	1	NODA_AZCCA	Q07739 azorhizobiu
31	39	40.2	264	1	KKA3_ENTFA	P00554 enterococcu
32	39	40.2	283	1	HXA9_FUGRU	Q42506 fugu rubrip
33	39	40.2	286	1	RIP1_MOMCH	P16094 momordica c

RESULT 1

ALIGNMENTS

Q8en66	oceanobacil	1	ILVC_OCEIH	344	40.2	39
Q9cph1	pasteurella	387	YBJF_PASMU	387	40.2	35
P05184	salmonella	415	PGTA_SALTY	415	40.2	36
O64203	mycobacteri	493	VG10_BPMD2	493	40.2	37
P16046	simian cyto	589	VP40_SCMVC	589	40.2	38
P52369	equine herp	643	VP40_HSVE2	643	40.2	39
Q9v128	thermotoga	688	DNLJ_THEMA	688	40.2	40
O9v128	pyrococcus	732	EF2_PYRAB	732	40.2	41
O59521	pyrococcus	732	EF2_PYRHO	732	40.2	42
P43158	porphyromon	868	PRTT_PORGI	868	40.2	43
P59418	rhizobium s	1102	TRAA_RHISN	1102	40.2	44
P55013	squalus aca	1191	S122_SQUAC	1191	40.2	45
Q55982	synechocyst	349	PDXA_SYNY3	349	39.7	46
P26827	thermoanaer	710	CDGT_THETU	710	39.7	47
Q911p8	bacterioph	184	VP50_BPAPS	184	39.2	48
Q98pr2	mycoplasma	197	RUVA_MYCPU	197	39.2	49
P74591	synechocyst	290	AROE_SYNY3	290	39.2	50
P26882	bos taurus	370	CYP4_BOVIN	370	39.2	51
Q08752	homo sapien	370	CYP4_HUMAN	370	39.2	52
Q9cr16	mus musculu	370	CYP4_MOUSE	370	39.2	53
Q28275	canis famil	522	FINC_CANTA	522	39.2	54
Q28377	equus cabal	522	FINC_HORSE	522	39.2	55
Q15822	homo sapien	640	PPCM_HUMAN	640	39.2	56
O60568	homo sapien	738	PLO3_HUMAN	738	39.2	57
Q8ygn1	brucella me	740	PURL_BRUME	740	39.2	58
Q90997	gallus gall	776	TFR1_CHICK	776	39.2	59
P58727	felis silve	833	TLR4_FELCA	833	39.2	60
P34261	caenorhabdi	838	YKAA_CAEEL	838	39.2	61
P07589	bos taurus	2265	FINC_BOVIN	2265	39.2	62
P11276	mus musculu	2477	FINC_MOUSE	2477	39.2	63
P3172	herpes simp	393	VGLD_HSV2	393	38.7	64
P44776	haemophilus	428	FUCP_HAEIN	428	38.7	65
Q9v264	homo sapien	503	AGP4_HUMAN	503	38.7	66
P27036	bacillus oh	704	CDGT_BACOH	704	38.7	67
P36000	saccharomyc	726	ADB2_YEAST	726	38.7	68
Q50704	mycobacteri	120	YY24_MYCTU	120	38.1	69
Q9pr82	ureaplasma	158	Y063_UREPA	158	38.1	70
Q9-w10	deinococcus	160	MENG_DEIRA	160	38.1	71
P04673	rhizobium l	196	NODA_RHILT	196	38.1	72
P04338	rhizobium l	196	NODA_RHILV	196	38.1	73
Q53252	rhizobium t	196	NODA_RHITR	196	38.1	74
Q8vvf0	sinorhizobi	197	NODA_SINTE	197	38.1	75
O93980	aspergillus	197	COX5_ASPNG	197	38.1	76
Q52839	rhizobium l	197	NODA_RHILO	197	38.1	77
P03253	mycoplasma	232	RS3_MYCCA	232	38.1	78
P46247	acholeplasm	241	RS3_ACHFL	241	38.1	79
Q94487	streptomyce	261	NAGB_STRCO	261	38.1	80
P00471	bacterioph	286	TYSY_BPT4	286	38.1	81
P46700	mycobacteri	296	THT2_MYCLE	296	38.1	82
P96888	mycobacteri	297	THT2_MYCTU	297	38.1	83
Q9a812	caulobacter	299	SYGA_CAUCR	299	38.1	84
O05252	bacillus su	350	YUFN_BACSU	350	38.1	85
Q866q5	escherichia	375	YBJF_ECO57	375	38.1	86
P75817	escherichia	375	YBJF_ECOLI	375	38.1	87
Q8ymg7	anabaena sp	384	H182_ANASP	384	38.1	88
P53891	saccharomyc	406	YNQ5_YEAST	406	38.1	89
P28724	giardia lam	449	DHE4_GIALA	449	38.1	90
YK4077	mycobacteri	452	YK97_MYCLE	452	38.1	91
Q9vsh2	drosophila	498	G56A_DROME	498	38.1	92
Q03263	saccharomyc	540	YH8M_YEAST	540	38.1	93
P49000	rattus norv	553	MIS_RAT	553	38.1	94
P27106	mus musculu	555	MIS_MOUSE	555	38.1	95
P03972	bos taurus	575	MIS_BOVIN	575	38.1	96
Q07307	emeritella	615	UAPA_EMENI	615	38.1	97
Q81020	human papil	633	VEI_HPVS4	633	38.1	98
P11229	bacterioph	648	VEI_BPPH6	648	38.1	99
P55577	rhizobium s	726	Y4NA_RHISN	726	38.1	100

```

YB_JF_SALTY
ID YB_JF_SALTY STANDARD; PRT; 375 AA.
AC Q82841;
DT 28-FEB-2003 (Rel. 41, Created)
DE Hypothetical RNA methyltransferase ybJF (EC 2.1.1.-).
GN YB_JF OR STV0915 OR T2014.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feitwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Could be a 23S rRNA (Uracil-5-)-methyltransferase.
CC -!- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. YB_JF
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627268; CAD05321.1; -.
CC HAMAP; MF_01012; -.
CC InterPro; IPR000051; SAM_bind.
CC PROSITE; PS01230; TRMA_1; 1.
CC PROSITE; PS01231; TRMA_2; 1.
CC KW Hypothetical protein; Transferase; Methyltransferase; rRNA processing;
CC FT ACT_SITE 334 BY SIMILARITY.
CC SQ SEQUENCE 375 AA; 42070 MW; 40C6A0563FA92598 CRC64;
Query Match 47.4%; Score 46; DB 1; Length 375;
Best Local Similarity 58.8%; Pred. No. 2.6;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
QY 1 NPTVA--LYGLKQDWEH 15
DB 214 NPTVASRLYATARDWVG 230
RESULT 2
YB_JF_SALTY STANDARD; PRT; 376 AA.
ID YB_JF_SALTY
AC Q82QJ5;
DT 28-FEB-2003 (Rel. 41, Created)
DE Hypothetical RNA methyltransferase ybJF (EC 2.1.1.-).
GN YB_JF OR STV0915 OR T2014.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Could be a 23S rRNA (Uracil-5-)-methyltransferase.
CC -!- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. YB_JF
CC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE008737; AAL19818.1; -.
CC HAMAP; MF_01012; ybJF.
CC InterPro; IPR000051; SAM_bind.
CC PROSITE; PS01230; TRMA_1; 1.
CC PROSITE; PS01231; TRMA_2; 1.
CC KW Hypothetical protein; Transferase; Methyltransferase; rRNA processing;
CC FT ACT_SITE 334 BY SIMILARITY.
CC SQ SEQUENCE 376 AA; 42180 MW; 343163D8E7FE0BFF CRC64;
Query Match 47.4%; Score 46; DB 1; Length 376;
Best Local Similarity 58.8%; Pred. No. 2.6;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
QY 1 NPTVA--LYGLKQDWEH 15
DB 214 NPTVASRLYATARDWVG 230
RESULT 3
UPK2_RHILO STANDARD; PRT; 264 AA.
ID UPK2_RHILO
AC Q98N11;
DT 28-FEB-2003 (Rel. 41, Created)
DE Hypothetical RNA methyltransferase ybJF (EC 2.1.1.-).
GN YB_JF OR STV0915 OR T2014.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,

```


Best Local Similarity 60.0%; Pred. No. 3.9;
Matches 9; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 1 NPTV--ALYGLKQDW 13
: | | | | | | | | | |
Db 40 HPARYVLYGKQDW 54

RESULT 6

ID MENG_YERPE STANDARD; PRT; 161 AA.
AC Q8ZJJ7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-adenosylmethionine:2-demethylmenaquinone methyltransferase
DE (EC 2.1.1.-)
GN MENG OR YPO0103 OR Y0292.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlysheva A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
RC -!- FUNCTION: Converts dimethylmenaquinone (DMK) to menaquinone (MK)
(By similarity).
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE MENG FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AJ141414; CAC88968.1; -;
DR EMBL; AF001328; AAM63864.1; -;
DR HAMAP; MF_00471; -; 1.
DR InterPro; IPR005493; Methyltransf_6.
DR Pfam; PF03737; Methyltransf_6; 1.
KW Menaquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 161 AA; 17315 MW; C325EF41BFEA938B CRC64;

Query Match
Best Local Similarity 43.3%; Score 42; DB 1; Length 161;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGI 16
: | | | | | | | | | |
Db 75 NAEALAEALKNEWEGI 90

RESULT 7

ID SL123_PSEAM STANDARD; PRT; 1023 AA.
AC P55019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Solute carrier family 12 member 3 (Thiazide-sensitive sodium-chloride
DE cotransporter) (Na-Cl symporter).
GN SLC12A3 OR TSC.
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
OS americanus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bladder urothelium;
RX MEDLINE=93219361; PubMed=8464884;
RA Gamba G., Saltzberg S.N., Lombardi M., Miyanoshta A., Lytton J.,
RA Hediger M.A., Brenner B.M., Hebert S.C.;
RT "Primary structure and functional expression of a cDNA encoding the
RT thiazide-sensitive, electroneutral sodium-chloride cotransporter."
RL Proc. Natl. Acad. Sci. U.S.A. 90:2749-2753(1993).
CC -!- FUNCTION: Electrically silent transporter system. Mediates sodium
CC and chloride reabsorption.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Urinary bladder.
CC -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L11615; AAA49272.1; -;
DR PIR; A47296; A47296.
DR InterPro; IPR002293; AA/rel_permeasel.
DR InterPro; IPR004842; KCL cotransp.
DR InterPro; IPR004841; Permease.
DR Pfam; PF00324; aa_permeases; 1.
DR TIGRFAMs; TIGR00930; 2a30; 1.

DR Transport; Ion transport; Sodium transport; Symport;
KW Transmembrane; 1
FT DOMAIN 1 132 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 133 133 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT DOMAIN 185 215 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 259 279 POTENTIAL.
FT DOMAIN 280 283 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 284 304 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT DOMAIN 358 374 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 452 472 POTENTIAL.
FT DOMAIN 473 510 POTENTIAL.
FT TRANSMEM 511 531 POTENTIAL.
FT TRANSMEM 577 597 POTENTIAL.
FT DOMAIN 598 745 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 746 766 POTENTIAL.
FT TRANSMEM 831 851 POTENTIAL.
FT DOMAIN 852 1023 CYTOPLASMIC (POTENTIAL).

```
FT DOMAIN 77 82 POLY-ARG.
FT DOMAIN 117 122 POLY-GLU.
FT DOMAIN 407 412 POLY-SER.
FT DOMAIN 933 937 POLY-GLN.
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 543 543 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1023 AA; 112340 MW; 69AE2D53B8F84D89 CRC64;

Query Match 43.3%; Score 42; DB 1; Length 1023;
Best Local Similarity 58.3%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDW 13
Db . 727 PNVLLMGFKDW 738

RESULT 8
RL11_SPOFR STANDARD; PRT; 195 AA.
AC Q962U2;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 60S ribosomal protein L11.
GN RPL11.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Amphipyridae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RA Landais I., Ogliastro M., Mita K., Nohata J., Lopez-Ferber M.,
RA Duonor-Cerutti M., Fournier P., Devauchelle G.;
RA "Full-length ribosomal protein sequence from an EST library of
RT Spodoptera frugiperda cells (Sf9).";
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSOCIATES WITH 5S RNA.
CC -!- SIMILARITY: BELONGS TO THE L5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF400182; AAK92154.1; -.
DR InterPro: IPR002132; Ribosomal_L5.
DR InterPro: IPR003236; Ribosomal_L5_mit.
DR Pfam: PF00281; Ribosomal_L5; 1.
DR Pfam: PF00673; Ribosomal_L5.C; 1.
DR PROSITE: PS00358; RIBOSOMAL_L5; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 195 AA; 22346 MW; A34A304552E1BFB5 CRC64;

Query Match 42.8%; Score 41.5; DB 1; Length 195;
Best Local Similarity 25.0%; Pred. No. 7.8;
Matches 8; Conservative 5; Mismatches 2; Indels 17; Gaps 1;

QY 1 NPTVALYGL-----KQDWEG 15
Db 135 DESIGYGLDFYVVLGPTRIOCTQKTQDWQG 166
|||||

RESULT 9
FLAV_ANASP STANDARD; PRT; 169 AA.
AC P11241;
```

```
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Flavodoxin.
GN ISIB OR ALR2405.
OS Anabaena sp. (strain PCC 7120), and
OS Anabaena sp. (strain PCC 7119).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690, 1168;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7120;
RX MEDLINE=89296496; PubMed=2500643;
RA Leonhardt K.G., Straus N.A.;
RT "Sequence of the flavodoxin gene from Anabaena variabilis 7120.";
RL Nucleic Acids Res. 17:4384-4384(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7120;
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7119;
RX MEDLINE=92074973; PubMed=1720613;
RA Fillat M.F., Borrias W.E., Weisbeek P.J.;
RT "Isolation and overexpression in Escherichia coli of the flavodoxin
RL gene from Anabaena PCC 7119.";
RN Biochem. J. 280:187-191(1991).
RN [4]
RP SEQUENCE OF 1-36.
RC STRAIN=PCC 7119;
RX MEDLINE=90381288; PubMed=2119231;
RA Fillat M.F., Edmondson D.E., Gomez-Moreno C.;
RT "Structural and chemical properties of a flavodoxin from Anabaena PCC
RT 7119.";
RL Biochim. Biophys. Acta 1040:301-307(1990).
RN [5]
RP STRUCTURE BY NMR.
RC STRAIN=PCC 7120;
RX MEDLINE=91104858; PubMed=2125478;
RA Stockman B.J., Krezel A.M., Markley J.L., Leonhardt K.G.,
RA Straus N.A.;
RT "Hydrogen-1, carbon-13, and nitrogen-15 NMR spectroscopy of Anabaena
RT 7120 flavodoxin: assignment of beta-sheet and flavin binding site
RT resonances and analysis of protein-flavin interactions.";
RL Biochemistry 29:9600-9609(1990).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RC STRAIN=PCC 7120;
RX MEDLINE=93271891; PubMed=1303762;
RA Rao S.T., Shaffie F., Yu C., Satyshur K.A., Stockman B.J.,
RA Markley J.L., Sundaralingam M.;
RT "Structure of the oxidized long-chain flavodoxin from Anabaena 7120
RT at 2-A resolution.";
RL Protein Sci. 1:1413-1427(1992).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
RC STRAIN=PCC 7120;
RA Burkhardt B.M., Ramakrishnan B., Yan H., Reedstrom R.J., Markley J.L.,
RA Straus N.A., Sundaralingam M.;
RT "Structure of the trigonal form of recombinant oxidized flavodoxin
RT from Anabaena 7120 at 1.40-A resolution.";
RL Acta Crystallogr. D 51:318-330(1995).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
```

STRAIN=PCC 7119;
MEDLINE=99318886; PubMed=10388575;
Fernandez-Reco J., Romero A., Sancho J.;
"Energetics of a hydrogen bond (charged and neutral) and of a
cation-pi interaction in apoflavodoxin.";
J. Mol. Biol. 290:319-330(1999).
-/- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX
ENZYMES.
-/- COFACTOR: FMN.
-/- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
-/- SIMILARITY: Contains 1 flavodoxin-like domain.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; S68006; AAB20462.1; -
EMBL; AP003589; BAB74104.1; -
EMBL; X14577; CAA32720.1; -
PIR; AF2106; AF2106.
PDB; 1FLV; 31-OCT-93.
PDB; 1FTG; 23-DEC-96.
PDB; 1RCG; 26-JAN-95.
PDB; 1QHE; 20-MAY-99.
PDB; 1DX9; 10-APR-00.
InterPro: IPR001226; Flavodoxin.
Pfam: PF002258; flavodoxin; 1.
PROSITE: PS00201; FLAVODOXIN; 1.
PROSITE: PS0902; FLAVODOXIN_LIKE; 1.
Electron transport; Flavoprotein; FMN; 3D-structure;
Complete proteome.
INIT_MET 0 0
DOMAIN 4 164 0
STRAND 4 8
HELIX 14 25
TURN 26 35
STRAND 31 35
TURN 36 37
HELIX 41 46
STRAND 49 53
TURN 56 57
STRAND 58 60
TURN 61 62
HELIX 64 70
TURN 71 72
HELIX 73 75
TURN 79 80
STRAND 82 88
TURN 91 96
TURN 98 99
HELIX 100 111
TURN 112 113
STRAND 115 116
TURN 120 121
STRAND 123 124
TURN 131 132
STRAND 133 134
TURN 135 136
STRAND 137 138
TURN 141 143
STRAND 145 147
TURN 149 151
HELIX 152 166
TURN 167 169
SQ SEQUENCE 169 AA; 18832 MW; BB0BBD0B6DEF9A58 CRC64;
Query Match 42.3%; Score 41; DB 1; Length 169;
Best Local Similarity 47.1%; Pred. No. 8.2;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
"

QY 2 PTVALYGLKQDWEGISS 18
|| : | : ||| : |
Db 55 PTWNIGELQSDWEGLYS 71

RESULT 10
ALR3_SALTI
ID ALR3_SALTI STANDARD: PRT; 368 AA.
AC Q82300;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alanine racemase 3 (EC 5.1.1.1).
GN ALR3 OR STY3763 OR T3513.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -/- FUNCTION: Provides the D-alanine required for cell wall
CC biosynthesis (By similarity).
CC -/- CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -/- COFACTOR: Pyridoxal phosphate (By similarity).
CC -/- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; first
CC step.
CC -/- SIMILARITY: Belongs to the alanine racemase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627279; CAD09518.1; -
CC EMBL; AE016846; AAC71021.1; -
CC HAMAP; MF_01201; -; 1.
CC InterPro: IPR000821; Ala racemase.
CC Pfam; PF00842; Ala racemase_C; 1.
CC Pfam; PF01168; Ala racemase_N; 1.
CC PRINTS; PR00992; ALARACEMASE.
CC PROSITE; PS00395; ALANINE_RACEMASE; 1.
CC Isomerase; Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis;
CC Complete proteome.
KW ACT_SITE 42 42 CATALYTIC BASE SPECIFIC TO D-ALANINE
FT ACT_SITE 262 262 (BY SIMILARITY).
FT ACT_SITE 262 262 CATALYTIC BASE SPECIFIC TO L-ALANINE
FT (BY SIMILARITY).

```
FT BINDING 42 42 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 368 AA: 397778 MW: 1338AGD2936004B3 CRC64;

Query Match 42.3%; Score 41; DB 1; Length 368;
Best Local Similarity 58.3%; Pred No. 19;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PTVALLYGLKQDW 13
| |||:| |
Db 222 PGVALFGVAQPW 233

RESULT 11
ID YK97_MYCTU STANDARD; PRT; 452 AA.
AC Q10706;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein RV2097c.
GN RV2097C OR MT2158 OR MTCY49.37C.
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch K., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Ufferback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: STRONG. TO MLEPRAE ML1328.
CC -!- SIMILARITY: TO RHODOCOCUS ERXTHROPOLIS HYPOTHETICAL PROTEIN IN
CC THCR 5'REGION (ORF6) (AC F43484).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z73966; CAA98209.1; -.
DR EMBL: AE007065; AAK46439.1; -.
DR PIR: D70768; D70768.
DR TIGR: MT2158; -.
DR Tuberculist: RV2097c; -.
DR InterPro: IPR004347; DUF245.
DR InterPro: IPR004989; DUF275.
DR Pfam: PF03136; DUF245; 1.
DR Pfam: PF03316; DUF275; 1.
KW Hypothetical protein; Complete proteome.
```

```
SQ SEQUENCE 452 AA: 51384 MW: 12AF8B7872D2F5EA CRC64;

Query Match 42.3%; Score 41; DB 1; Length 452;
Best Local Similarity 47.6%; Pred No. 23;
Matches 10; Conservative 1; Mismatches 6; Indels 4; Gaps 1;

QY 2 PTVALLYGLKQD-----WEGISS 18
| | | | | | | | | | | | | |
Db 161 PKAATYCLSQRAEHIEGVSS 181

RESULT 12
MTHC_DROME
ID MTHC_DROME STANDARD; PRT; 475 AA.
AC P83119;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable G-protein-coupled receptor Mth-like 12 precursor (Methuselah-
DE like 12 protein).
GN MTHL12 OR MTH-LIKE-12.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Chery J.M., Cui Y., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fustin C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reintner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP IDENTIFICATION.
RX MEDLINE=21173629; PubMed=11274391;
RA West A.P. Jr., Llamas L.L., Snow P.M., Benzer S., Bjorkman P.J.;
RT "Crystal structure of the ectodomain of Methuselah, a Drosophila G
```

```

RT  protein-coupled receptor associated with extended lifespan."
RL  Proc. Natl. Acad. Sci. U.S.A. 98:3744-3749(2001).
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED
CC  RECEPTORS. MTH SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE003699; -; NOT_ANNOTATED_CDS.
DR  FLYBASE; FBgn0045442; mchl12.
DR  GO; GO:0004930; P:G-protein coupled receptor activity; ISS.
DR  GO; GO:0008340; P:determination of adult life span; ISS.
DR  GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; ISS.
DR  GO; GO:0006950; P:response to stress; ISS.
DR  InterPro; IPR000832; GPCR_secretin.
DR  Pfam; PF00002; 7tm2; 1.
DR  PROSITE; PS00649; G_PROTEIN_RECP_F2_1; FALSE_NEG.
DR  PROSITE; PS00650; G_PROTEIN_RECP_F2_2; FALSE_NEG.
DR  PROSITE; PS0261; G_PROTEIN_RECP_F2_4; 1.
KW  Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein;
KW  Signal; Multigene family.
FT  SIGNAL 1 17
FT  CHAIN 18 475
FT  DOMAIN 18 214
FT  TRANSMEM 215 235
FT  DOMAIN 236 242
FT  TRANSMEM 243 263
FT  DOMAIN 264 296
FT  TRANSMEM 297 307
FT  DOMAIN 308 328
FT  TRANSMEM 329 360
FT  DOMAIN 361 381
FT  TRANSMEM 382 403
FT  DOMAIN 404 424
FT  TRANSMEM 425 442
FT  DOMAIN 443 463
FT  TRANSMEM 464 475
FT  DISULFID 27 81
FT  DISULFID 83 88
FT  DISULFID 92 183
FT  DISULFID 93 104
FT  DISULFID 149 203
FT  CARBOHYD 19 19
FT  CARBOHYD 34 34
FT  CARBOHYD 55 55
FT  CARBOHYD 135 135
FT  CARBOHYD 352 352
SQ  SEQUENCE 475 AA; 55424 MW; 3590FEF87F77264 CRC64;

Query Match 42.3%; Score 41; DB 1; Length 475;
Best Local Similarity 46.2%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDWE 14
Db 322 PTWVIFSMNQWE 334

RESULT 13
PYRG_AZOBZ STANDARD; PRT; 544 AA:
AC P28595:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

protein-coupled receptor associated with extended lifespan).
PYRG.
GN Azospirillum brasilense.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Azospirillum.
OX NCBI_TaxID=192;
(1)
RN SEQUENCE FROM N.A.
RP STRAIN=Sp7 / ATCC 29145;
RX MEDLINE=94186025; Pubmed=8138139;
RA Zimmer W.; Hundersagen B.;
RT Identification and sequencing of pyrG, the CTP synthetase gene of
RT Azospirillum brasilense Sp7.;
RL FEMS Microbiol. Lett. 115:273-278(1994).
CC -!- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
CC either L-glutamine or ammonia as the source of nitrogen.
CC -!- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -!- ENZYME REGULATION: Allosterically activated by GTP, when glutamine
CC is the substrate. Inhibited by CTP.
CC -!- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
CC (last) step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the CTP synthase family.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL; X67216; CAA47656.1; -
DR  PIR; I39496; S25101.
DR  HAMAP; MF_01227; -; 1.
DR  InterPro; IPR000991; GATase_1.
DR  InterPro; IPR004468; PyrG.
DR  Pfam; PF00117; GATase; 1.
DR  TIGRFAMs; TIGR00337; PyrG; 1.
DR  PROSITE; PS00442; GATASE_TYPE_I; 1.
KW  Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase.
FT  DOMAIN 1 300
FT  DOMAIN 301 544
FT  ACT_SITE 380 380
FT  ACT_SITE 516 516
FT  ACT_SITE 518 518
FT  SEQUENCE 544 AA; 59908 MW; 13AA1A8B8EB6C5DB CRC64;

Query Match 42.3%; Score 41; DB 1; Length 544;
Best Local Similarity 53.6%; Pred. No. 28;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDW 13
Db 410 NPVVGLGLMTW 422

RESULT 14
PLOS_MOUSE
ID PLOS_MOUSE STANDARD; PRT; 741 AA.
AC Q9R0E1; O9CY9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 precursor
DE (EC 1.14.11.4) (Lysyl hydroxylase 3) (LH3).
GN PLOS3.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
(1)

```

RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=99357020; PubMed=10429951;
 RA Ruotsalainen H., Sipila L., Kerkela E., Pospiech H., Myllylae R.;
 RT "Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3,
 their phylogenetic analysis and tissue-specific expression in the
 mouse";
 RL Matrix Biol. 18:325-329(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SVJ;
 RX MEDLINE=21233587; PubMed=11334715;
 RA Ruotsalainen H., Vanhatupa S., Tampio M., Sipila L., Valtavaara M.,
 RA Myllylae R.;
 RT "Complete genomic structure of mouse lysyl hydroxylase 2 and lysyl
 hydroxylase 3/collagen glucosyltransferase";
 RL Matrix Biol. 20:137-146(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Glissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -!- FUNCTION: FORMS HYDROXYLYSINE RESIDUES IN -XAA-LYS-GLY- SEQUENCES
 IN COLLAGENS. THESE HYDROXYLYSINES SERVE AS SITES OF ATTACHMENT
 FOR CARBOHYDRATE UNITS AND ARE ESSENTIAL FOR THE STABILITY OF THE
 INTERMOLECULAR COLLAGEN CROSSLINKS.
 CC -!- CATALYTIC ACTIVITY: Procollagen L-lysine + 2-oxoglutarate + O(2) -
 procollagen 5-hydroxy-L-lysine + succinate + CO(2).
 CC -!- COFACTOR: IRON AND ASCORBATE.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: MEMBRANE BOUND IN CISTERNAE OF ROUGH
 ENDOPLASMIC RETICULUM.
 CC -!- TISSUE SPECIFICITY: Highly expressed in the heart, lung, liver and
 testis.
 CC -!- SIMILARITY: BELONGS TO THE LYSYL HYDROXYLASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF046783; AAD54618.1; -;
 DR EMBL; AY014830; AAK00576.1; -;
 DR EMBL; AK031315; BAB28704.1; -;
 DR MGD; MGI:1347008; Plod3
 DR InterPro; IPR005123; 2OG-FeII_Oxy
 DR InterPro; IPR006620; Pro_4_hyd_alph.
 DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
 DR ProDom; PD011578; ProcollLys_dioxy; 1.
 DR SMART; SM00702; P4HG; 1.
 DR PROSITE; PS01325; LYS_HYDROXYLASE; 1.

KW Oxidoreductase; Dioxigenase; Signal; Iron; Vitamin C; Glycoprotein;
 KW Endoplasmic reticulum; Membrane; Polymorphism.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 741 PROCOLLAGEN-LYSINE, 2-OXOGLUTARATE 5-
 FT DIOXYGENASE 3.
 FT METAL 670 670 IRON (BY SIMILARITY).
 FT METAL 672 672 IRON (BY SIMILARITY).
 FT METAL 722 722 IRON (BY SIMILARITY).
 FT ACT_SITE 732 732 POTENTIAL.
 FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 8 P -> H (IN REF. 3).
 SQ SEQUENCE 741 AA; 84922 MW; D1B79B386339DF4 CRC64;
 Query Match 42.3%; Score 41; DB 1; Length 741;
 Best Local Similarity 53.3%; Pred. No. 39;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWEG 15
 Db 66 NYTVRTLGLGQEWG 80
 ID QAL5_NEUCR STANDARD; PRT; 918 AA.
 AC P11637;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-FEB-1991 (Rel. 17, Last annotation update)
 DE Quinate repressor.
 GN QA-1S.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RX MEDLINE=89293848; PubMed=2525625;
 RA Geever R.F., Hulet L., Baum J.A., Tyler B.M., Patel V.B.,
 RA Rutledge B.J., Case M.E., Giles N.H.;
 RT "DNA sequence, organization and regulation of the qa gene cluster of
 Neurospora crassa";
 RL J. Mol. Biol. 207:15-34(1989).
 CC -!- FUNCTION: REPRESSOR FOR ENZYMES AND PROTEINS OF QUINATE
 METABOLISM.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X14603; CAA32753.1; -;
 DR PIR; S04255; S04255.
 DR InterPro; IPR001381; DkQuinase_I.
 DR InterPro; IPR000623; Shik_kinase.
 DR InterPro; IPR006151; Shikimate_DH.
 DR Pfam; PF01488; DkQuinase_I; 1.
 DR Pfam; PF01202; SKI; 1.
 DR Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
 SQ SEQUENCE 918 AA; 100580 MW; 67EDA399CBF098B2 CRC64;
 Query Match 42.3%; Score 41; DB 1; Length 918;
 Best Local Similarity 66.7%; Pred. No. 49;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 5 ALYGLKQDWEG 16

```

Db      684 ALYGTNDWIGI 695
|||||
DHPI1_SCHPO STANDARD; PRT; 991 AA.
AC      P40848;
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Protein dhp1.
GN      DHPI1 OR SPAC26A3 12C.
OS      Schizosaccharomyces pombe (Fission yeast).
OC      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC      Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OC      Schizosaccharomycetes.
OX      NCBI_TaxId=4896;
RN      [1]
RN      SEQUENCE FROM N.A., AND SEQUENCE OF 1-29.
RC      STRAIN=975;
RX      MEDLINE=94247347; PubMed=8190062;
RX      Sugano S., Shobuiki T., Takeda T., Sugino A., Ikeda H.;
RT      "Molecular analysis of the dhp1+ gene of Schizosaccharomyces pombe:
RT      an essential gene that has homology to the DST2 and RAT1 genes of
RT      Saccharomyces cerevisiae.";
RL      Mol. Gen. Genet. 243:1-8(1994).
RN      [2]
RN      SEQUENCE FROM N.A.
RC      STRAIN=972;
RX      MEDLINE=2184401; PubMed=11859360;
RX      Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA      Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA      Brooks K., Brown D., Brown S., Chillingworth J., Churcher C.M.,
RA      Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA      Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA      Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA      James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA      Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odel C.,
RA      Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA      Rutherford J., Rutter S., Saunders R., Seeger K., Sharp S.,
RA      Skelton J., Simmonds M., Squares R., Squares K., Stevens K.,
RA      Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA      Woodward J., Volckart G., Aert R., Robben J., Grymonprez B.,
RA      Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA      Gabel C., Fuchs M., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA      Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA      Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA      Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA      Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA      Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA      Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA      Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA      Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT      "The genome sequence of Schizosaccharomyces pombe.";
RL      Nature 415:871-880(2002).
RN      [3]
RN      CHARACTERIZATION.
RX      MEDLINE=21138436; PubMed=11238999;
RX      Shobuiki T., Tatebayashi K., Tani T., Sugano S., Ikeda H.;
RT      "The dhp1+ gene, encoding a putative nuclear 5'3' exonuclease, is
RT      required for proper chromosome segregation in fission yeast.";
RL      Nucleic Acids Res. 29:1326-1333(2001).
CC      -!- FUNCTION: Essential for vegetative cell growth; required for
CC      proper chromosome segregation. Can bind to DNA and has 5'->3'
CC      exonuclease activity.
CC      -!- SUBUNIT: Interacts with dhp1.
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- SIMILARITY: Belongs to the 5'-3' exonuclease family.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D17752; BAA04601.1; -.
DR      MBL; Z69240; CAA93235.1; -.
DR      PIR; S43891; S43891.
DR      GenBank; SPAC26A3.12C; -.
DR      InterPro; IPR004859; Put_53exo.
DR      Pfam; PF03159; XRN_N_1.
KW      Nuclear protein; Hydrolase; Nuclease; Exonuclease.
FT      DOMAIN 264 288
SQ      SEQUENCE 991 AA; 112367 MW; 83FA34D93DADAD00 CRC64;
Query Match 42.3%; Score 41; DB 1; Length 991;
Best Local Similarity 41.2%; Pred. No. 53;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY      1 NPTVALYGLKQDWEGIS 17
Db      706 NPTIDLNGKKFEWQGVA 722
RESULT 17
FLAV_SYNTP7 STANDARD; PRT; 169 AA.
AC      P10340;
DT      01-MAR-1989 (Rel. 10, Created)
DT      01-JAN-1990 (Rel. 13, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Flavodoxin.
GN      ISIB.
OS      Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC      Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX      NCBI_TaxId=1140;
RN      [1]
RN      SEQUENCE FROM N.A.
RX      MEDLINE=88086879; PubMed=3121586;
RT      "Isolation, sequence analysis, and transcriptional studies of the
RT      flavodoxin gene from Anacystis nidulans R2.";
RL      J. Bacteriol. 170:258-265(1988).
RN      [2]
RN      SEQUENCE OF 1-55, AND X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX      MEDLINE=83216115; PubMed=6406674;
RA      Smith W.W., Pattidge K.A., Ludwig M.L., Petsko G.A., Tsernoglou D.,
RA      Tanaka M., Yasunobu K.T.;
RT      "Structure of oxidized flavodoxin from Anacystis nidulans.";
RL      J. Mol. Biol. 165:737-755(1983).
RN      [3]
RN      X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX      MEDLINE=20079529; PubMed=10610791;
RA      Drennan C.L., Pattidge K.A., Weber C.H., Metzger A.L., Osborne C.,
RA      Hoover D.M., Drennan C.L., Metzger A.L., Osborne C., Weber C.H.,
RA      Ludwig M.L.;
RT      "Refined structures of oxidized flavodoxin from Anacystis nidulans.";
RL      J. Mol. Biol. 294:711-724(1999).
RN      [4]
RN      X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RX      MEDLINE=20079530; PubMed=10610792;
RA      Hoover D.M., Drennan C.L., Metzger A.L., Osborne C., Weber C.H.,
RA      Pattidge K.A., Ludwig M.L.;
RT      "Comparisons of wild-type and mutant flavodoxins from Anacystis
RT      nidulans. Structural determinants of the redox potentials.";
RL      J. Mol. Biol. 294:725-743(1999).
RN      [5]
RN      STRUCTURE BY NMR.
RX      MEDLINE=91329335; PubMed=1907844;
RA      Clubb R.T., Thanabal V., Osborne C., Wagner G.;
RT      "1H and 15N resonance assignments of oxidized flavodoxin from
RT      Anacystis nidulans with 3D NMR.";
RL      Biochemistry 30:7718-7730(1991).

```

```
CC -!- FUNCTION: LOW-POTENTIAL ELECTRON DONOR TO A NUMBER OF REDOX
CC ENZYMES.
CC -!- COFACTOR: FMN.
CC -!- INDUCTION: BY iron stress.
CC -!- SIMILARITY: BELONGS TO THE FLAVODOXIN FAMILY.
CC -!- SIMILARITY: Contains 1 flavodoxin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19116; AAA22050.1; -.
DR PDB; 1OFV; 31-JAN-94.
DR PDB; 1CZH; 29-DEC-99.
DR PDB; 1CZK; 29-DEC-99.
DR PDB; 1CZL; 29-DEC-99.
DR PDB; 1CZN; 29-DEC-99.
DR PDB; 1CZO; 29-DEC-99.
DR PDB; 1CZR; 29-DEC-99.
DR PDB; 1CZU; 29-DEC-99.
DR PDB; 1D03; 29-DEC-99.
DR PDB; 1D04; 29-DEC-99.
DR InterPro; IPR001226; Flavodoxin.
DR Pfam; PF00258; flavodoxin; 1.
DR PROSITE; PS00201; FLAVODOXIN; 1.
DR PROSITE; PS00902; FLAVODOXIN-LIKE; 1.
KW Electron transport; Flavoprotein; FMN; 3D-structure.
FT INIT_MET 0 0
FT DOMAIN 3 164 FLAVODOXIN-LIKE.
FT CONFLICT 54 54 C -> S (IN REF. 2).
FT CONFLICT 167 168 FG -> GF (IN REF. 2).
FT STRAND 3 7
FT HELIX 13 25
FT TURN 26 26
FT TURN 28 30
FT STRAND 31 35
FT HELIX 36 38
FT HELIX 41 46
FT STRAND 49 53
FT TURN 56 57
FT TURN 58 60
FT STRAND 61 62
FT HELIX 64 69
FT HELIX 70 75
FT TURN 79 80
FT STRAND 82 88
FT TURN 91 96
FT TURN 98 99
FT HELIX 100 111
FT TURN 112 113
FT STRAND 115 116
FT STRAND 120 121
FT TURN 123 124
FT TURN 131 132
FT STRAND 133 134
FT TURN 135 136
FT STRAND 137 138
FT STRAND 141 143
FT TURN 145 147
FT HELIX 149 166
FT TURN 167 169
SQ SEQUENCE 169 AA; 18646 MW; CF049E12F9CGAAIF CRC64;

Query Match 41.2%; Score 40; DB 1; Length 169;
Best Local Similarity 53.3%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 PTVALYGLKQDWEGI 16
|| : | : |||||
```

```
Db 55 PTWNGELQSDWEGI 69

RESULT 18
NODA_MESPL STANDARD; PRT; 196 AA.
ID NODA_MESPL STANDARD; PRT; 196 AA.
AC Q8VVH1; Q8VVH2;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nodulation protein A (EC 2.3.1.-).
GN NODA.
OS Mesorhizobium plurifarum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=69974;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORS1096, and ORS1255;
RA Ba S., Willems A., Lorquin J., Roche P., de Lajudie P., Neyra M.,
RA Moulin L., Gillis M., Dreyfus B., Boivin-Masson C.;
RT "Symbiotic and taxonomic diversity of Acacia tortilis rhizobia.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: N-acetyltransferase required for nodulation. Acts in the
CC production of a small, heat-stable compound (Nod) that stimulates
CC mitosis in various plant protoplasts.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the noda family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ302678; CAC80541.1; -.
DR EMBL; AJ302680; CAC80543.1; -.
DR HAMAP; MF_00084; -.
DR InterPro; IPR003484; Noda.
DR Pfam; PF02474; Noda; 1.
DR PRODOM; PD004579; Noda; 1.
DR PROSITE; PS01349; NODA; 1.
KW Transferase; Acyltransferase; Nodulation.
FT VARIANT 8 8 R -> K (in strain ORS1096).
FT VARIANT 16 16 Q -> E (in strain ORS1096).
FT VARIANT 21 21 I -> V (in strain ORS1096).
FT VARIANT 69 69 M -> I (in strain ORS1096).
FT VARIANT 77 77 K -> R (in strain ORS1096).
FT VARIANT 102 102 I -> L (in strain ORS1096).
FT VARIANT 106 106 L -> F (in strain ORS1096).
FT VARIANT 109 110 VR -> LL (in strain ORS1096).
FT VARIANT 114 114 Q -> R (in strain ORS1096).
FT VARIANT 155 156 RD -> HF (in strain ORS1096).
FT VARIANT 174 174 L -> V (in strain ORS1096).
SQ SEQUENCE 196 AA; 21707 MW; D1BDA45DAACE3850 CRC64;

Query Match 41.2%; Score 40; DB 1; Length 196;
Best Local Similarity 46.2%; Pred. No. 14;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 4 VALYGLKQDWEGI 16
: |||::: |||
Db 88 LGLYGIKRDLEGL 100

RESULT 19
PYRF_BACSU STANDARD; PRT; 239 AA.
ID PYRF_BACSU STANDARD; PRT; 239 AA.
AC P25971;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
```

DR	Subtilist; BG10719; pyzF.
DR	HAMAP; MF_01200; -; 1.
DR	InterPro: IPR001754; OMPdecase.
DR	Pfam: PF00215; OMPdecase; 1.
DR	PROSITE: PS00156; OMPDECASE; 1.
KW	Lysase; Decarboxylase; Pyrimidine biosynthesis; 3D-structure;
KW	Complete proteome.
FT	ACT_SITE 62
SQ	SEQUENCE 239 AA; 25992 MW; DB1743714ED052E7 CRC64; 1 NPTVALYGLKODWEGISS 18 PROTON DONOR. 222 DPVKAYKAVRLEWESIKS 239 I : I :: :IIII : : : :IIII
QY	Query Match 41.2%; Score 40; DB 1; Length 239; Best Local Similarity 38.9%; Pred. No. 18; Matches 7; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
DB	222 DPVKAYKAVRLEWESIKS 239
RESULT 20	
ID	Y33B_MYCPN STANDARD; PRT; 341 AA.
AC	P75302; 30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hypothetical protein MG335.2 homolog (P01_orf341).
GN	MPN483 OR MP359.
OS	Mycoplasma pneumoniae.
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX	NCBI_TaxID=2104;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-ATCC 29342 / M129;
RX	MEDLINE-97105885; PubMed-8948633;
RA	Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA	Herrmann R.;
RT	"Complete sequence analysis of the genome of the bacterium Mycoplasma
RL	pneumoniae.";
RL	Nucleic Acids Res. 24:4420-4449(1996).
CC	-!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC	STRONG, TO M.GENITALIUM MG335.2.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AE000034; AAB96007.1; -.
DR	PIR; S73685; S73685
DR	InterPro: IPR001173; Glyco.trans.2.
DR	Pfam: PF00535; Glycos.transf.2; 1.
KW	Hypothetical protein; Transferase; Glycosyltransferase;
KW	Complete proteome.
SQ	SEQUENCE 341 AA; 40414 MW; C209F50D714CB3D0 CRC64; Query Match 41.2%; Score 40; DB 1; Length 341; Best Local Similarity 38.9%; Pred. No. 26; Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	9 LKQDWEGI 16
DB	26 LKQDWNGV 33
RESULT 21	
ID	GLG2_RHIME STANDARD; PRT; 486 AA.
AC	P58394;
DT	28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Glycogen synthase 2 (EC 2.4.1.21) (Starch [bacterial glycogen]
 synthase 2).
 GN GIGA2 OR RBL411 OR SMB20704.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymb (megaplasmid 2).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396508; PubMed=11481431;
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
 RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie J.,
 RA Golding B., Puehler A.;
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
 RT fixing endosymbiont Sinorhizobium meliloti.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894 (2001).
 CC -!- FUNCTION: Synthesizes alpha-1,4-glucan chains using ADP-glucose.
 CC ADP + {(1,4)-alpha-D-glucosyl}(N+1).
 CC -!- PATHWAY: Glycogen biosynthesis; second step.
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 1.
 CC Bacterial/plant glycogen synthase subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; AL603647; CAC49811.1; -;
 CC FIRM; C96018; C96018.
 CC HAMAP; MF_00484; -; 1.
 CC InterPro; IPR001296; Glyco_transf_1.
 CC Pfam; PF0534; Glycos_transf_1.
 CC Glycogen biosynthesis; Transferase; Glycosyltransferase; Plasmid;
 KW Complete proteome.
 FT BINDING 15 15 ADP-GLUCOSE (BY SIMILARITY).
 SQ SEQUENCE 486 AA; 53635 MW; 0068BAF9E8FE805 CRC64;

 Query Match 41.2%; Score 40; DB 1; Length 486;
 Best Local Similarity 50.0%; Pred. No. 37;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWEGI 16
 DB 437 TFLYLRRLRVWEG 450

 RESULT 22
 PDAT_YEAST
 ID PDAT_YEAST STANDARD; PRT; 661 AA.
 AC P40345;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phospholipid:diacylglycerol acyltransferase (EC 2.3.1.158) (PDAT).
 GN LRO1 OR YNR008W OR N2042.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / FY1679;
 RX MEDLINE=95208356; PubMed=7900425;
 RA Verhaaselt P., Aert R., Voet M., Volckaert G.;
 RT "Twelve open reading frames revealed in the 23.6 kb segment flanking

the centromere on the Saccharomyces cerevisiae chromosome XIV right
 arm.";
 RL Yeast 10:1355-1361 (1994).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=20300927; PubMed=10829075;
 RA Dahlqvist A., Stahl U., Lenman M., Banas A., Lee M., Sandager L.,
 RA Ronne H., Styenne S.;
 RT "Phospholipid:diacylglycerol acyltransferase; an enzyme that catalyzes
 RT the acyl-CoA-independent formation of triacylglycerol in yeast and
 RT plants";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:6487-6492 (2000).
 CC -!- FUNCTION: Triacylglycerol formation by an acyl-CoA independent
 CC pathway. The enzyme specifically transfers acyl groups from the
 CC sn-2 position of a phospholipid to diacylglycerol, thus forming
 CC an sn-1-lysophospholipid.
 CC -!- CATALYTIC ACTIVITY: Phospholipid + 1,2-diacylglycerol =
 CC lysophospholipid + triacylglycerol.
 CC -!- SIMILARITY: SOME, TO MAMMALIAN PHOSPHATIDYLCHOLINE-STEROL O-
 CC ACYLTRANSFERASE.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; X77395; CAA54576.1; -;
 CC EMBL; Z71623; CAA96285.1; -;
 CC PIR; S45131; S45131.
 CC SGD; S0005291; LRO1.
 CC GO; GO:0046027; F:phospholipid:diacylglycerol acyltransferase. ; IDA.
 CC GO; GO:0019915; P:lipid storage; IDA.
 CC GO; GO:0019432; P:triacylglycerol biosynthesis; IDA.
 CC InterPro; IPR003386; LACT.
 CC Pfam; PF02450; LACT; 1.
 CC Transferase; Acyltransferase; Transmembrane.
 KW TRANSMEM 81 101
 FT TRANSFERASE
 SQ SEQUENCE 661 AA; 75393 MW; 01C043319A836F44 CRC64;

 Query Match 41.2%; Score 40; DB 1; Length 661;
 Best Local Similarity 35.7%; Pred. No. 52;
 Matches 10; Conservative 3; Mismatches 3; Indels 12; Gaps 1;
 QY 3 TVALYGLK-----QDWEGISS 18
 DB 389 TLAMYGLEKFFSRIERVKMLQTWGGIPS 416

 RESULT 23
 PURL_ZYMMO
 ID PURL_ZYMMO STANDARD; PRT; 734 AA.
 AC Q9REQ6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (PGAM
 DE synthase II).
 GN PUR-Q.
 OS Zymomonas mobilis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
 OC Sphingomonadaceae; Zymomonas.
 OX NCBI_TaxID=542;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 31821 / ZM4 / CP4;
 RA Um H.W., Kang H.S.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
 CC ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
 CC (formamido)-N(1)-(5-phospho-D-ribosyl)acetamidine + L-glutamate.

CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
 CC -1- SUBUNIT: Heterodimer of two subunits, purQ and purL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF213822; AAF23789.1; --
 CC HAMAB; MF_00420; -- 1.
 CC InterPro: IPR000728; AIRS_related.
 CC Pfam; PF00586; AIRS; 2.
 CC Pfam; PF02769; AIRS_C; 2.
 CC Purine biosynthesis; Ligase; ATP-binding.
 CC NP_BIND 106 117 ATP (POTENTIAL).
 CC SQ SEQUENCE 734 AA; 17679 MW; B771635E0F66A166 CRC64;
 CC -----
 CC Query Match 41.2%; Score 40; DB 1; Length 734;
 CC Best Local Similarity 58.3%; Pred. No. 58;
 CC Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 CC -----
 CC QY 2 PTVALYGLKQDW 13
 CC II: II III
 CC Db 550 PTIGVGLLQDW 561
 CC -----
 CC RESULT 24
 CC NKCL_MANSE STANDARD; PRT: 1060 AA.
 CC ID Q25479; 1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE humetanide-sensitive sodium-(Potassium)-chloride cotransporter
 CC (NA-K-CL symporter).
 CC OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 CC OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Spinglodes;
 CC OC Sphingidae; Sphinginae; Manduca.
 CC OX NCBI_TaxID=71130;
 CC RN [1]
 CC SEQUENCE FROM N.A.
 CC RC TISSUE=Malpighian tubules;
 CC RX MEDLINE=96035837; PubMed=7550244;
 CC RA Reagan J.D.;
 CC RT "Molecular cloning of a putative Na(+)-K(+)-2Cl-cotransporter from
 CC the Malpighian tubules of the tobacco hornworm, Manduca sexta.";
 CC RL Insect Biochem. Mol. Biol. 25:875-880(1995).
 CC -1- FUNCTION: Electrically silent transporter system. Mediates sodium
 CC and chloride reabsorption. Plays a vital role in the regulation of
 CC ionic balance and cell volume.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U17344; AAA75600.1; --
 CC PIR; T30823; T30823.
 CC InterPro: IPR002293; AA/rel_permease1.
 CC DR InterPro: IPR004842; KCL cotransp.
 CC DR InterPro: IPR002443; NaKCL cotransp.
 CC PRINTS; PR01207; NAKCLFNSPRT.

DR TIGREMS; TIGR00930; 2a30; 1.
 KW Transport; Ion transport; Sodium transport; Symport;
 KW Potassium transport; Potassium; Transmembrane.
 FT DOMAIN 1 122 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 123 143 POTENTIAL.
 FT TRANSMEM 154 174 POTENTIAL.
 FT DOMAIN 175 197 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 198 218 POTENTIAL.
 FT TRANSMEM 230 270 POTENTIAL.
 FT DOMAIN 271 275 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 276 296 POTENTIAL.
 FT TRANSMEM 332 352 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 353 367 POTENTIAL.
 FT TRANSMEM 368 388 POTENTIAL.
 FT TRANSMEM 432 452 POTENTIAL.
 FT DOMAIN 453 497 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 498 518 POTENTIAL.
 FT TRANSMEM 563 583 POTENTIAL.
 FT DOMAIN 584 642 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 643 663 POTENTIAL.
 FT TRANSMEM 882 902 POTENTIAL.
 FT DOMAIN 903 1060 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 419 419 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 816 816 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1060 AA; 117392 MW; 4C39D1F0B645FEFC CRC64;
 CC -----
 CC Query Match 41.2%; Score 40; DB 1; Length 1060;
 CC Best Local Similarity 58.3%; Pred. No. 85;
 CC Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 CC -----
 CC QY 2 PTVALYGLKQDW 13
 CC II: II III
 CC Db 719 PNVLMLGYKSDW 730
 CC -----
 CC RESULT 25
 CC DDX8_HUMAN STANDARD; PRT: 1220 AA.
 CC ID DDX8_HUMAN
 CC AC Q14562;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE ATP-dependent helicase DDX8 (RNA helicase HRH1) (DEAH-box protein 8).
 CC GN DDX8.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC OX NCBI_TaxID=9606;
 CC RN [1]_TaxID=9606;
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=95021291; PubMed=7935475;
 CC RA Ono Y.; Ono M.; Shimura Y.;
 CC RT "Identification of a putative RNA helicase (HRH1), a human homolog of
 CC yeast Prp22.";
 CC RL Mol. Cell. Biol. 14:7611-7620(1994).
 CC RN [2]
 CC RP FUNCTION.
 CC RX MEDLINE=96196415; PubMed=8608946;
 CC RA Ono M.; Shimura Y.;
 CC RT "A human RNA helicase-like protein, HRH1, facilitates nuclear export
 CC of spliced mRNA by releasing the RNA from the spliceosome.";
 CC RL Genes Dev. 10:997-1007(1996).
 CC -1- FUNCTION: FACILITATES NUCLEAR EXPORT OF SPLICED MRNA BY RELEASING
 CC THE RNA FROM THE SPLICESOME.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DOMAIN: THE RS DOMAIN CONFERS A NUCLEAR LOCALIZATION SIGNAL, AND
 CC APPEARS TO FACILITATE THE INTERACTION WITH THE SPLICESOME.
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEADH
 CC SUBFAMILY. DDX8/PRP22 ORTHOLOG.
 CC -1- SIMILARITY: Contains 1 S1 motif domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: D50487; BAA09078.1; -
 DR PIR: A56236; A56236.
 DR HSP: P05055; LSRO.
 DR Genew: HGNC:2749; DDX8.
 DR GK: Q14562; -
 DR MIM: 600396; -
 DR GO: GO:0005681; C:spliceosome complex; TAS.
 DR GO: GO:0004004; F:ATP dependent RNA helicase activity; TAS.
 DR GO: GO:0008248; F:pre-mRNA splicing factor activity; TAS.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD_box.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR003029; S1.
 DR Pfam: PF04408; HA2; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00575; S1; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00316; S1; 1.
 DR PROSITE: PS00690; DEAD_ATP_HELICASE; 1.
 DR PROSITE: PS01266; S1; 1.
 KW mRNA processing; mRNA splicing; Helicase; ATP-binding;
 KW Nuclear protein.
 FT DOMAIN 172 175 POLY-LYS.
 FT DOMAIN 176 228 ARG/SER-RICH (RS DOMAIN).
 FT DOMAIN 265 336 S1 MOTIF.
 FT NP_BIND 588 595 ATP (POTENTIAL).
 FT SITE 685 688 DEAD BOX.
 FT MUTAGEN 594 594 K->E: IN GET; INHIBITION OF PRE-MRNA
 FT MUTAGEN 717 717 S->L: IN LAT; INHIBITION OF PRE-MRNA
 FT SPlicing AND NUCLEAR EXPORT OF UNSPLICED
 FT RNA.
 FT SEQUENCE 1220 AA; 139314 MW; 17C1602A73A0EF24 CRC64;

Query Match 41.2%; Score 40; DB 1; Length 1220;
 Best Local Similarity 53.8%; Pred. No. 98;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 VALYGLKQDWEGI 16
 | | | | | | | |
 Db 282 VQLEGLRKRWEGL 294

RESULT 26
 PKHD_HUMAN STANDARD; PRT; 4074 AA.
 AC Q8TCZ9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Polycystic kidney and hepatic disease 1 precursor (Fibrocystin)
 DE (Polyductin) (Tigmin).
 GN PKHD1 OR FCYT OR TIGM1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 1), VARIANTS ARPKD MET-36; VAL-222;
 RP TRP-1249; ARG-1917; GLY-1995; LYS-2331;
 RP THR-2957; PHE-3018 AND THR-3553, AND VARIANTS MET-752; CYS-760;
 RP ARG-852; VAL-1262; MET-2938; TYR-3139; ILE-3960 AND ARG-4048.
 RC TISSUE-Kidney;

RX MEDLINE-21918597; PubMed-11919560;
 RA Ward C.J., Hogan M.C., Rossetti S., Walker D., Sneddon T., Wang X.,
 RA Kubly V., Cunningham J.M., Bacallao R., Ishibashi M., Milliner D.S.,
 RA Torres V.E., Harris P.C.;
 RT "The gene mutated in autosomal recessive polycystic kidney disease
 RT encodes a large, receptor-like protein.";
 RL Nat. Genet. 30:259-269(2002).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING, AND VARIANTS
 RP ARPKD MET-36; VAL-222; LEU-253; HIS-760; SER-1122; TRP-1624; THR-2957
 RP AND TYR-3139.
 RC TISSUE-Kidney;
 RX MEDLINE-21947447; PubMed-11898128;
 RA Onuchic L.F., Furu L., Nagasawa Y., Hou X., Eggemann T., Ren Z.,
 RA Bergmann C., Senderek J., Esquivel E., Zeltner R.,
 RA Rudnik-Schoenborn S., Mrug M., Sweeney W., Avner E.D., Zerres K.,
 RA Guay-Woodford L.M., Somlo S., Germino G.G.;
 RT "PKHD1, the polycystic kidney and hepatic disease 1 gene, encodes a
 RT novel large protein containing multiple
 RT immunoglobulin-like-plexin-transcription-factor domains and parallel
 RT beta-helix 1 repeats.";
 RL Am. J. Hum. Genet. 70:1305-1317(2002).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT TYR-1204.
 RC TISSUE-Fetal kidney;
 RX MEDLINE-22074934; PubMed-12079288;
 RA Xiong H., Chen Y., Yi Y., Tsuchiya K., Moekel G., Cheung J.,
 RA Liang D., Tham K., Xu X., Chen X.-Z., Pei Y., Zhao Z.J., Wu G.;
 RT "A novel gene encoding a TIG multiple domain protein is a positional
 RT candidate for autosomal recessive polycystic kidney disease.";
 RL Genomics 80:96-104(2002).
 CC -!- FUNCTION: May be a receptor protein that acts in collecting-duct
 CC and biliary differentiation.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms-2;
 CC Comment-Additional isoforms seem to exist. As a matter of fact,
 CC alternatively spliced products seem to fall into two broad
 CC groups: one group, which includes the longest continuous ORF but
 CC which may also include molecules lacking some middle domains,
 CC has a single TM element and is likely to be associated with the
 CC plasma membrane. The other group lacks a TM domain and thus its
 CC members may be secreted;
 CC Name=1;
 CC IsoId=Q8TCZ9-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q8TCZ9-2; Sequence=VSP_003947, VSP_003948;
 CC -!- TISSUE SPECIFICITY: Predominantly expressed in fetal and adult
 CC kidney. Also present in the adult pancreas, but at much lower
 CC levels. Detectable in fetal and adult liver. Rather indistinct
 CC signal in fetal brain.
 CC -!- DISEASE: Defects in PKHD1 are the cause of all typical forms of
 CC autosomal recessive polycystic kidney disease (ARPKD), a
 CC hereditary and severe form of polycystic kidney disease affecting
 CC the kidneys and the hepatic biliary tract. The clinical spectrum
 CC is widely variable, with most cases presenting during infancy. The
 CC fetal phenotypic features classically include enlarged and
 CC echogenic kidneys, as well as oligohydramnios secondary to a poor
 CC urine output. Up to 50% of the affected neonates die shortly after
 CC birth, as a result of severe pulmonary hypoplasia and secondary
 CC respiratory insufficiency. In the subset that survives the
 CC perinatal period, morbidity and mortality are mainly related to
 CC severe systemic hypertension, renal insufficiency, and portal
 CC hypertension due to portal-tract fibrosis.
 CC -!- SIMILARITY: Contains 9 PBH1 repeats.
 CC -!- SIMILARITY: Contains 12 TIG/IPT domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```

RT jannaschli."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- PATHWAY: Arginine biosynthesis.
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (by
CC similarity).
CC -1- SIMILARITY: Belongs to the cara family.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67544; AAB99021.1; -.
DR HSSP: P00907; IC50.
DR TIGR: M01019; -.
DR HAMAP: MF_01209; -.
DR InterPro: IPR006220; Anth_synth1.
DR InterPro: IPR006274; Cara_small.
DR InterPro: IPR001317; CPS_GATase.
DR InterPro: IPR002474; CPSase_sm_chain.
DR InterPro: IPR000991; GATase_1.
DR Pfam: PF00988; CPSase_sm_chain; 1.
DR Pfam: PF00117; GATase; 1.
DR PRINTS: PR00097; ANTSNTASE11.
DR PRINTS: PR00099; CPSGATASE.
DR PRINTS: PR00096; GATASE.
DR TIGRPFAM: TIGR01368; CPSase1small; 1.
DR PROSITE: PS00442; GATASE_TYPE_1; FALSE_NEG.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase;
KW Glutamine amidotransferase; Complete proteome.
FT DOMAIN 1 171
FT DOMAIN 172 354
FT ACT_SITE 246 246 GLUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 246 246 GATASE (BY SIMILARITY).
SQ SEQUENCE 354 AA; 39920 MW; 4D8F07776CA65F9F CRC64;

Query Match 40.7%; Score 39.5; DB 1; Length 354;
Best Local Similarity 40.0%; Pred. No. 32;
Matches 8; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 2 PTVALYGLKQDM---EGISS 18
DB 56 PLEGNVGAKKDFESDGKA 75

RESULT 28
ID NODA_RHIS3 STANDARD; PRT; 196 AA.
AC P72329;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nodulation protein A (EC 2.3.1.-).
GN NODA.
OS Rhizobium sp. (strain N33).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=103798;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96303537; PubMed=8755627;
RA Cloutier J., Laberge S., Prevost D., Antoun H.;
RT Rhizobium sp. (Oxytropis arctobia) strain N33, a nitrogen-fixing
RT microsybiont of both arctic and temperate legumes."

```

```

RL Mol. Microbe Interact. 9:523-531(1996).
CC -1- FUNCTION: N-ACYLTRANSFERASE REQUIRED FOR NODULATION. ACTS IN THE
CC PRODUCTION OF A SMALL, HEAT-STABLE COMPOUND (NOD) THAT STIMULATES
CC MITOSIS IN VARIOUS PLANT PROTOPLASTS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the noda family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U53327; AAB16892.1; -.
DR HAMAP: MF_00084; -.
DR InterPro: IPR003484; Noda.
DR Pfam: PF02474; Noda; 1.
DR ProDom: PD004579; Noda; 1.
DR PROSITE: PS01349; NODA; 1.
KW Transferase; Acyltransferase; Nodulation.
SQ SEQUENCE 196 AA; 21727 MW; CA065D2046F3F061 CRC64;

Query Match 40.2%; Score 39; DB 1; Length 196;
Best Local Similarity 46.2%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 VALYGLKQDMEGI 16
DB 88 IGLYGVRPDLGG 100

RESULT 29
ID RS3_OCEIH STANDARD; PRT; 213 AA.
AC P59182;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S3.
GN RPS3 OR OB0125.
OS Oceanobacillus theyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTEB31 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus theyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RT Nucleic Acids Res. 30:3927-3935(2002).
CC -1- FUNCTION: Binds the lower part of the 30S subunit head. Binds mRNA
CC in the 70S ribosome, positioning it for translation (by
CC similarity).
CC -1- SUBUNIT: Part of the 30S ribosomal subunit. Forms a tight complex
CC with proteins S10 and S14 (by similarity).
CC -1- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.
CC -1- SIMILARITY: Contains 1 KH type-2 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP004593; BAC12081.1; -.
DR HAMAP: MF_01309; -.
DR SMART: SM00322; KH, 1.

```

DR TIGR01009; tpsc_bact; 1.
DR PROSITE; PS05823; KH_type_2; 1.
DR PROSITE; PS00548; RIBOSOMAL_S3; 1.
KW Ribosomal protein; RNA-binding; Complete proteome.
FT DOMAIN 38 106 KH type-2.
SQ SEQUENCE 213 AA; 23534 MW; 3543002B6C3B2934 CRC64;

Query Match 40.2%; Score 39; DB 1; Length 213;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 NPTVALYGLKQME 14
||| |
Db 6 NPTGLRVGIKQME 19

RESULT 30
NODA_AZOCA STANDARD; PRT; 226 AA.
AC 007739;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nodulation protein A (EC 2.3.1.-).
GN NODA.
OS Acetivibrium caulinodans.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OX Hypnomicrobacteraceae; Azorhizobium.
RN NCBI_TaxID=7;
RP SEQUENCE FROM N.A.
RC MEDLINE=90136519; PubMed=2615763;
RA Goethals K., Gao M., Tomekpe K., van Montagu M., Holsters M.;
RT "Common nodABC genes in Nod locus 1 of Azorhizobium caulinodans:
RT nucleotide sequence and plant-inducible expression.";
RL Mol. Gen. Genet. 219:289-299(1989).
CC -1- FUNCTION: N-ACETYLTRANSFERASE REQUIRED FOR NODULATION. ACTS IN THE
CC PRODUCTION OF A SMALL, HEAT-STABLE COMPOUND (NOD) THAT STIMULATES
CC MITOSIS IN VARIOUS PLANT PROTOPLASTS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the noda family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L18897; AAB51162.1; -
DR PIR; J00393; J00393.
DR HAMAP; ME_00084; -; 1.
DR InterPro; IPR003484; Noda.
DR Pfam; PF02474; Noda; 1.
DR ProDom; PD004579; Noda; 1.
DR PROSITE; PS01349; Noda; 1.
KW Transference; Acyltransferase; Nodulation.
SQ SEQUENCE 226 AA; 24915 MW; F1992B421A002315 CRC64;

Query Match 40.2%; Score 39; DB 1; Length 226;
Best Local Similarity 46.2%; Pred. No. 25;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 4 VALYGLKQMEGI 16
|:|:|:|:|:|
Db 118 LGLYGVRLDEGL 130

RESULT 31
KKA3_ENTFA STANDARD; PRT; 264 AA.
ID KKA3_ENTFA

AC P00554;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aminoglycoside 3'-phosphotransferase (EC 2.7.1.95) (Kanamycin kinase,
DE type III) (Neomycin kanamycin phosphotransferase, type III)
DE (Aph(3')III).
GN APH.
OS Enterococcus faecalis (Streptococcus faecalis), and
OS Staphylococcus aureus.
OC Plasmid pJH1
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351, 1280;
RN (1)
RP SEQUENCE FROM N.A.
RC SPECIES-E. faecalis;
RX MEDLINE=84029883; PubMed=6313476;
RA Tricu-Cuot P., Courvillain P.;
RT "Nucleotide sequence of the streptococcus faecalis plasmid gene
RT encoding the 3'/5'-aminoglycoside phosphotransferase type III.";
RL Gene 23:331-341(1983).
[2]

RP SEQUENCE FROM N.A.
RC SPECIES-S. aureus;
RX MEDLINE=88174299; PubMed=6100986;
RA Gray G.S., Fitch W.M.;
RT "Evolution of antibiotic resistance genes: the DNA sequence of a
RT kanamycin resistance gene from Staphylococcus aureus.";
RL Mol. Biol. Evol. 1:57-66(1983).
CC -1- FUNCTION: RESISTANCE TO KANAMYCIN AND STRUCTURALLY RELATED
CC AMINOGLYCOSIDES, INCLUDING AMIKACIN.
CC -1- CATALYTIC ACTIVITY: ATP + kanamycin = ADP + kanamycin 3'-
CC phosphate.
CC -1- SIMILARITY: TO OTHER AMINOGLYCOSIDE PHOSPHOTRANSFERASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01547; CAA24789.1; -
DR PIR; M6771; AAA26596.1; ALT_SEQ.
DR EMBL; X99398; CAA67773.1; -
DR PIR; A06655; PKSOJF.
DR PIR; A26220; PKSAF.
DR PDB; 1J7I; 08-AUG-01.
DR PDB; 1J7U; 08-AUG-01.
DR PDB; 1J7U; 08-AUG-01.
DR PDB; 1L8T; 15-JUN-02.
DR PDB; 1L8U; 19-JUN-02.
DR InterPro; IPR002575; APH.
DR Pfam; PF01636; APH; 1.
KW Antibiotic resistance; Transferase; Kinase; ATP-binding; Plasmid;
KW 3D-structure.
FT VARIANTS 190 190 BY SIMILARITY.
FT VARIANTS 35 35 MISSING (IN S. AUREUS).
SQ SEQUENCE 264 AA; 30974 MW; 7EBA5851D380C5A CRC64;

Query Match 40.2%; Score 39; DB 1; Length 264;
Best Local Similarity 63.6%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 6 LYGLKQMEGI 16
|:|:|:|:|:|
Db 244 LGIKPDMEXI 254

RESULT 32
HXA9_FUGRU STANDARD; PRT; 283 AA.
ID HXA9_FUGRU

```

AC 042506;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A9 (Hoxa-9).
GN HOXA9.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97285126; PubMed=9140399;
RA Aparicio S., Hawker K., Cottage A., Mikawa Y., Zuo L., Venkatesh B.,
RA Chen E., Krumlauf R., Brenner S.;
RT "Organization of the Fugu rubripes Hox clusters: evidence for
RT continuing evolution of vertebrate Hox complexes.";
RL Nat. Genet. 16:79-83(1997).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE ABD-B HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U92573; AAB68684.1; -.
DR HSSP: P02833; 9ANT.
DR TRANSFAC: T03772; -.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR006711; Hox9_act.
DR InterPro: IPR000047; HTH_Lambdaressr.
DR Pfam: PF00046; homeobox; 1.
DR Pfam: PF04617; Hox9_act; 1.
DR PRINTS: PR00031; HTHREPRESSR.
DR PRODOM: PD000010; Homeobox; 1.
DR SMART: SM00389; HOX; 1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DNA BIND 216 275 HOMEBOX.
SQ SEQUENCE 283 AA; 31410 MW; FDD3D052702BE36C CRC64;

Query Match 40.2%; Score 39; DB 1; Length 283;
Best Local Similarity 47.1%; Pred. No. 31;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 PYVALYGLKQDMEGISS 18
   11 11:1 1:1
Db 118 PTHHYGIKPDGIGVRS 134

RESULT 33
RIP1_MOMCH STANDARD; PRT; 286 AA.
ID RIP1_MOMCH
AC P16094; P24697;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribosome-inactivating protein momordin I precursor (rRNA
DE N-glycosidase) (EC 3.2.2.22) (Alpha-momorcharin) (Alpha-MMC).
OS Momordica charantia (Bitter melon) (Balsam pear).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

DR Glycosylated; P16094; -.
 DR InterPro: IPR001574; RIP.
 DR Pfam: PF00161; RIP; 1.
 DR PRINTS: PR00396; SHIGARICIN.
 DR PROSITE: PS00275; SHIGA_RICIN; 1.
 DR Plant defense; Protein synthesis inhibitor; Hydrolase; Toxin; Signal;
 KM Glycoprotein; 3D-structure.
 FT SIGNAL 1 23
 FT CHAIN 24 269
 FT PROPEP 270 286
 FT ACT_SITE 183 183
 FT CARBOND 250 250
 FT STRAND 25 28
 FT HELIX 34 47
 FT TURN 54 57
 FT STRAND 57 60
 FT HELIX 66 66
 FT STRAND 70 76
 FT TURN 78 78
 FT STRAND 82 88
 FT TURN 89 92
 FT STRAND 93 96
 FT TURN 100 101
 FT STRAND 102 105
 FT HELIX 109 114
 FT TURN 115 117
 FT TURN 120 121
 FT STRAND 124 127
 FT HELIX 134 141
 FT HELIX 145 147
 FT STRAND 150 150
 FT HELIX 152 162
 FT TURN 163 163
 FT HELIX 167 167
 FT STRAND 187 187
 FT HELIX 188 195
 FT TURN 196 197
 FT STRAND 202 202
 FT HELIX 206 225
 FT TURN 226 230
 FT STRAND 231 238
 FT TURN 240 241
 FT STRAND 246 250
 FT TURN 251 252
 FT HELIX 254 258
 FT TURN 259 259
 FT STRAND 260 260
 FT STRAND 263 263
 FT HELIX 266 268
 SQ SEQUENCE 286 AA; 31532 MW; E1H013ABEBC216CF CRC64;
 Query Match 40.2%; Score 39; DB 1; Length 286;
 Best Local Similarity 37.5%; Pred. No. 32;
 Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
 QY 2 PTVALLYGLKQDMEG15 17
 DB 204 PSLATISLENSWSG15 219
 ID ILVC_OCEIH STANDARD; PRT; 344 AA.
 AC Q8EN66;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
 isomeroreductase) (Alpha-keto-beta-hydroxyacyl reductoisomerase).
 GN ILVC OR ORB2621.
 OS Oceanobacillus theyensis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
 OX NCBI_TaxID=182710;
 RX (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-HTE831 / DSM 14371 / JCM 11309;
 RX MEDLINE-22220767; PubMed-12235376;
 RA Takami H., Takaki Y., Uchiyama I.;
 RT "Genome sequence of Oceanobacillus theyensis isolated from the Iheya
 Ridge and its unexpected adaptive capabilities to extreme
 environments.";
 RL Nucleic Acids Res. 30:3927-3935(2002).
 CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
 CC = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
 CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
 CC -1- SIMILARITY: Belongs to the ketol-acid reductoisomerase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AP004601; BAC14577.1; -
 DR HAMAP: MF_00435; -; 1.
 DR InterPro: IPR000506; ACh_Isomrdctase.
 DR Pfam: PF01450; ILVC; 1.
 DR TIGRFAMs: TIGR00465; ILVC; 1.
 KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
 KW Complete proteome.
 FT ACT_SITE 107 107
 SQ SEQUENCE 344 AA; 37883 MW; 467BF2E67EB8E660 CRC64;
 Query Match 40.2%; Score 39; DB 1; Length 344;
 Best Local Similarity 63.6%; Pred. No. 38;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 5 ALYGLKQDMEG 15
 DB 148 ALYGVHDDYTG 158
 ID YBVF_PASMU STANDARD; PRT; 387 AA.
 AC Q8OCEH1
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
 isomeroreductase) (Alpha-keto-beta-hydroxyacyl reductoisomerase).
 GN YBVF_PASMU
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellales; Pasteurellaceae; Pasteurella.
 OX NCBI_TaxID=747;
 RX (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-Pm70;
 RX MEDLINE-21145866; PubMed-11248100;
 RA May B.J., Zhang O., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc Natl Acad Sci U S A. 98:3460-3465(2001).
 CC -1- FUNCTION: Could be a 23S rRNA (Uracil-5) methyltransferase.
 CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. YBVF
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@isd-sib.ch).
CC -----
DR EMBL: AE006042; AK02154.1; ALT_INT.
DR HAMAP: MF_01012; -: 1.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001566; TrmA.
DR PROSITE: PS01230; TRMA_1; 1.
DR PROSITE: PS01231; TRMA_2; 1.
KW Hypothetical protein; Transferase; Methyltransferase; rRNA processing;
KW Complete proteome.
FT ACI_SITE 346
FT ACI_SITE 346
SQ SEQUENCE 387 AA; 43876 MW; 81397AF98D943BE CRC64;
    BY SIMILARITY.

Query Match          40.2%; Score 39; DB 1; Length 387;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 9; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 1 NPPTVA--LYGLKQDW 13
   || || || || ||
Db 213 NPOVAGLYGTAAQHW 227

Search completed: October 6, 2003, 07:44:14
Job time : 7.10448 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 10.4776 Seconds
(without alignments)
72.688 Million cell updates/sec

Title: US-09-765-739A-6

Perfect score: 97

Sequence: 1 NPTVALXGLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	278	3	US-08-953-326-16
2	97	100.0	278	4	US-09-660-587-13
3	97	100.0	278	4	US-09-261-358A-13
4	97	100.0	278	4	US-09-201-458-9
5	97	100.0	278	4	US-09-314-701-10
6	88	90.7	280	4	US-09-660-587-11
7	88	90.7	280	4	US-09-261-358A-11
8	88	90.7	280	4	US-09-201-458-7
9	88	90.7	280	4	US-09-314-701-6
10	62	63.9	280	3	US-08-953-326-17
11	59	60.8	280	4	US-09-660-587-14
12	59	60.8	280	4	US-09-261-358A-14
13	59	60.8	280	4	US-09-201-458-10
14	59	60.8	280	4	US-09-314-701-12
15	59	60.8	280	4	US-09-314-701-32
16	58	59.8	276	3	US-08-953-326-18
17	58	59.8	280	3	US-08-733-230-4
18	58	59.8	280	3	US-08-953-326-4
19	58	59.8	281	4	US-09-660-587-9
20	58	59.8	281	4	US-09-261-358A-9
21	58	59.8	281	4	US-09-201-458-5
22	58	59.8	281	4	US-09-314-701-2
23	56	57.7	280	4	US-09-660-587-6
24	56	57.7	280	4	US-09-261-358A-6
25	56	57.7	280	4	US-09-314-701-38
26	56	57.7	284	4	US-09-660-587-15
27	56	57.7	284	4	US-09-261-358A-15

ALIGNMENTS

```
RESULT 1
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Njika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match          100.0%; Score 97; DB 3; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NPTVALYGLKQDWEGISS 18
DB      60 NPTVALYGLKQDWEGISS 77

RESULT 2
US-09-660-587-13
; Sequence 13, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-660-587-13

Query Match          100.0%; Score 97; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NPTVALYGLKQDWEGISS 18
DB      60 NPTVALYGLKQDWEGISS 77

RESULT 3
US-09-261-358A-13
; Sequence 13, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-261-358A-13

Query Match          100.0%; Score 97; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NPTVALYGLKQDWEGISS 18
DB      60 NPTVALYGLKQDWEGISS 77

RESULT 4
US-09-201-458-9
; Sequence 9, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-201-458-9

Query Match          100.0%; Score 97; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.8e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NPTVALYGLKQDWEGISS 18
DB      60 NPTVALYGLKQDWEGISS 77

RESULT 5
US-09-314-701-10
; Sequence 10, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
```

; APPLICANT: Ohasi, No. 654451710
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-10

Query Match 100.0%; Score 97; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.8e-09; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0;

QY 1 NPTVALYGLKQDWEGISS 18
Db 60 NPTVALYGLKQDWEGISS 77
|||||

RESULT 6
US-09-660-587-11
; Sequence 11, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-660-587-11

Query Match 90.7%; Score 88; DB 4; Length 280;
Best Local Similarity 83.3%; Pred. No. 2.1e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
Db 60 NPTVALYGLKQDWNGVSA 77
|||||

RESULT 7
US-09-261-358A-11
; Sequence 11, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 11

; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-261-358A-11

Query Match 90.7%; Score 88; DB 4; Length 280;
Best Local Similarity 83.3%; Pred. No. 2.1e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
Db 60 NPTVALYGLKQDWNGVSA 77
|||||

RESULT 8
US-09-201-458-7
; Sequence 7, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 7
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-201-458-7

Query Match 90.7%; Score 88; DB 4; Length 280;
Best Local Similarity 83.3%; Pred. No. 2.1e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
Db 60 NPTVALYGLKQDWNGVSA 77
|||||

RESULT 9
US-09-314-701-6
; Sequence 6, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 654451710
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-6

Query Match 90.7%; Score 88; DB 4; Length 280;
Best Local Similarity 83.3%; Pred. No. 2.1e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
Db 60 NPTVALYGLKQDWNGVSA 77
|||||

Db 60 NPTVALYGLKQDWDG 77

RESULT 10

US-08-953-326-17

; Sequence 17, Application US/08953326

; Patent No. 6251872

; GENERAL INFORMATION:

; APPLICANT: Barbet, Anthony F.

; APPLICANT: Ganta, Roman R.

; APPLICANT: McGuire, Travis C.

; APPLICANT: Burridge, Michael J.

; APPLICANT: Nyika, Aceme

; APPLICANT: Rurangirwa, Fred R.

; APPLICANT: Mahan, Suman M.

; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of

; FILE REFERENCE: UP-167C1

; CURRENT APPLICATION NUMBER: US/08/953,326

; EARLIER FILING DATE: 1997-10-17

; EARLIER FILING DATE: 1997-10-17

; EARLIER FILING DATE: 1997-10-17

; EARLIER FILING DATE: 1996-10-17

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 17

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis

US-08-953-326-17

Query Match 63.9%; Score 62; DB 3; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.0068; 2; Indels 0;

Matches 9; Conservative 4; Mismatches 2; Gaps 0;

Qy 1 NPTVALYGLKQDWDG 15

Db 60 NTTGVFGLKQDWDG 74

RESULT 11

US-09-660-587-14

; Sequence 14, Application US/09660587

; Patent No. 6392023

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: McBride, Jere W.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein

; FILE REFERENCE: D6152CIP2

; CURRENT APPLICATION NUMBER: US/09/660,587

; CURRENT FILING DATE: 2000-09-12

; PRIOR FILING DATE: 1999-03-03

; NUMBER OF SEQ ID NOS: 46

; SEQ ID NO 14

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis

; FEATURE:

; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F

US-09-660-587-14

Query Match 60.8%; Score 59; DB 4; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.022;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWDG 15

Db 60 NTTGVFGLKQDWDG 74

RESULT 12

US-09-261-358A-14

; Sequence 14, Application US/09261358A

; Patent No. 6403780

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: McBride, Jere W.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein

; FILE REFERENCE: D6152CIP

; CURRENT APPLICATION NUMBER: US/09/261,358A

; CURRENT FILING DATE: 1999-03-03

; PRIOR FILING DATE: 1998-11-30

; NUMBER OF SEQ ID NOS: 33

; SEQ ID NO 14

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis

; FEATURE:

; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F

US-09-261-358A-14

Query Match 60.8%; Score 59; DB 4; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.022;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWDG 15

Db 60 NTTGVFGLKQDWDG 74

RESULT 13

US-09-201-458-10

; Sequence 10, Application US/09201458A

; Patent No. 6458942

; GENERAL INFORMATION:

; APPLICANT: Walker, David H.

; APPLICANT: McBride, Jere W.

; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia

; FILE REFERENCE: D6152

; CURRENT APPLICATION NUMBER: US/09/201,458A

; CURRENT FILING DATE: 1998-11-30

; NUMBER OF SEQ ID NOS: 21

; SEQ ID NO 10

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis

; FEATURE:

; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F

US-09-201-458-10

Query Match 60.8%; Score 59; DB 4; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.022;

Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NPTVALYGLKQDWDG 15

Db 60 NTTGVFGLKQDWDG 74

RESULT 14

US-09-314-701-12

; Sequence 12, Application US/09314701

; Patent No. 6544517

; GENERAL INFORMATION:

; APPLICANT: Rikihisa, Yasuko

; APPLICANT: Ohasi, No. 6544517io

; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia

;; TITLE OF INVENTION: Chaffeensis
;; FILE REFERENCE: 22727/04021
;; CURRENT APPLICATION NUMBER: US/09/314,701
;; CURRENT FILING DATE: 1999-05-19
;; NUMBER OF SEQ ID NOS: 66
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 12
;; LENGTH: 280
;; TYPE: PRT
;; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-12

Query Match 60.8%; Score 59; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.022;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 60 NTTGVFGLKQDWG 74

RESULT 15
US-09-314-701-32
;; Sequence 32, Application US/09314701
;; Patent No. 6544517
;; GENERAL INFORMATION:
;; APPLICANT: Rikihisa, Yasuko
;; APPLICANT: Ohasi, No. 6544517io
;; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
;; FILE REFERENCE: 22727/04021
;; CURRENT APPLICATION NUMBER: US/09/314,701
;; CURRENT FILING DATE: 1999-05-19
;; NUMBER OF SEQ ID NOS: 66
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 32
;; LENGTH: 288
;; TYPE: PRT
;; ORGANISM: Ehrlichia canis
US-09-314-701-32

Query Match 60.8%; Score 59; DB 4; Length 288;
Best Local Similarity 60.0%; Pred. No. 0.023;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 60 NTTGVFGLKQDWG 74

RESULT 16
US-08-953-326-18
;; Sequence 18, Application US/08953326
;; Patent No. 6251872
;; GENERAL INFORMATION:
;; APPLICANT: Barbet, Anthony F.
;; APPLICANT: Ganta, Roman R.
;; APPLICANT: McGuire, Travis C.
;; APPLICANT: Burridge, Michael J.
;; APPLICANT: Nyika, Aceme
;; APPLICANT: Rurangirwa, Fred R.
;; APPLICANT: Mahan, Suman M.
;; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
;; FILE REFERENCE: UF-167C1
;; CURRENT APPLICATION NUMBER: US/08/953,326
;; CURRENT FILING DATE: 1997-10-17
;; EARLIER APPLICATION NUMBER: 08/953,326
;; EARLIER FILING DATE: 1997-10-17
;; EARLIER APPLICATION NUMBER: 08/733,230
;; EARLIER FILING DATE: 1996-10-17
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: PatentIn Ver. 2.0

;; SEQ ID NO 18
;; LENGTH: 276
;; TYPE: PRT
;; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match 59.8%; Score 58; DB 3; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.033;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NTTGVFGLKQDWG 73

RESULT 17
US-08-733-230-4
;; Sequence 4, Application US/08733230
;; Patent No. 6025338
;; GENERAL INFORMATION:
;; APPLICANT: Barbet, Anthony F.
;; APPLICANT: Ganta, Roman Reddy
;; APPLICANT: McGuire, Travis C.
;; APPLICANT: Burridge, Michael J.
;; APPLICANT: Nyika, Aceme
;; APPLICANT: Rurangirwa, Fred R.
;; APPLICANT: Mahan, Suman M.
;; TITLE OF INVENTION: Nucleic Acid Vaccines Against
;; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Saliwanchik & Saliwanchik
;; STREET: 2421 N.W. 41st Street, Suite A-1
;; CITY: Gainesville
;; STATE: FL
;; COUNTRY: USA
;; ZIP: 32606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/733,230
;; FILING DATE:
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Whitlock, Ted W.
;; REGISTRATION NUMBER: 36,965
;; REFERENCE/DOCKET NUMBER: UF-167
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 352-375-8100
;; TELEFAX: 352-372-5800
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 280 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-733-230-4

Query Match 59.8%; Score 58; DB 3; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.033;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 60 NTTGVFGLKQDWG 74

RESULT 18
US-08-953-326-4
;; Sequence 4, Application US/08953326

; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R. C.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match 59.8%; Score 58; DB 3; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.033;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 60 NTTGVFGLKQNDG 74

RESULT 19

US-09-660-587-9
; Sequence 9, Application US/09660587
; Patent No. 632023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-660-587-9

Query Match 59.8%; Score 58; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.034;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NTTGVFGLKQNDG 73

RESULT 20

US-09-261-358A-9
; Sequence 9, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:

; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-261-358A-9

Query Match 59.8%; Score 58; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.034;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NTTGVFGLKQNDG 73

RESULT 21

US-09-201-458-5
; Sequence 5, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xue-Jie
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-201-458-5

Query Match 59.8%; Score 58; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.034;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NTTGVFGLKQNDG 73

RESULT 22

US-09-314-701-2
; Sequence 2, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-2

Query Match      59.8%; Score 58; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.034;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NSTVGVLKQDWG 73

RESULT 23
US-09-660-587-6
; Sequence 6, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-6 protein
US-09-660-587-6

Query Match      57.7%; Score 56; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.074;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NSTVGVLKQDWG 73

RESULT 24
US-09-261-358A-6
; Sequence 6, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28SA3 protein
US-09-261-358A-6

Query Match      57.7%; Score 56; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.074;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NSTVGVLKQDWG 73

RESULT 25
US-09-314-701-38
; Sequence 38, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-38

Query Match      57.7%; Score 56; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.074;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
Db 59 NSTVGVLKQDWG 73

RESULT 26
US-09-660-587-15
; Sequence 15, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-660-587-15

Query Match      57.7%; Score 56; DB 4; Length 284;
Best Local Similarity 50.0%; Pred. No. 0.076;
Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
Db 63 TKAVEGLKKDWDGVKT 78

RESULT 27
US-09-261-358A-15
; Sequence 15, Application US/09261358A
```

```
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-261-358A-15

Query Match      57.7%; Score 56; DB 4; Length 284;
Best Local Similarity 50.0%; Pred. No. 0.076;
Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY      3 TVALYGLKQDWEGISS 18
Db      63 TKAVFGLKKDWDGKVT 78

RESULT 28
US-09-201-458-11
; Sequence 11, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 11
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-201-458-11

Query Match      57.7%; Score 56; DB 4; Length 284;
Best Local Similarity 50.0%; Pred. No. 0.076;
Matches 8; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY      3 TVALYGLKQDWEGISS 18
Db      63 TKAVFGLKKDWDGKVT 78

RESULT 29
US-09-660-587-2
; Sequence 2, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
```

```
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p38-7 protein
US-09-660-587-2

Query Match      54.6%; Score 53; DB 4; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.24;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      3 TVALYGLKQDWEG 15
Db      62 TVGVFGLKHDWDG 74

RESULT 30
US-09-261-358A-2
; Sequence 2, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28-1 protein
US-09-261-358A-2

Query Match      54.6%; Score 53; DB 4; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.24;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      3 TVALYGLKQDWEG 15
Db      62 TVGVFGLKHDWDG 74

RESULT 31
US-09-201-458-2
; Sequence 2, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
```

```
; OTHER INFORMATION: amino acid sequence of a 30 kDa immunoreactive
; OTHER INFORMATION: protein of Ehrlichia canis
US-09-201-458-2

Query Match          54.6%; Score 53; DB 4; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.24; Mismatches 2; Indels 0; Gaps 0;
Matches 8; Conservative 3;

QY 3 TVALYGLKQDWEG 15
Db 62 TVGVFLKHDWDG 74

RESULT 32
US-09-314-701-36
; Sequence 36, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Onasi, No. 6544517/lo
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-36

Query Match          54.6%; Score 53; DB 4; Length 307;
Best Local Similarity 61.5%; Pred. No. 0.27; Mismatches 2; Indels 0; Gaps 0;
Matches 8; Conservative 3;

QY 3 TVALYGLKQDWEG 15
Db 91 TVGVFLKHDWDG 103

RESULT 33
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match          53.6%; Score 52; DB 3; Length 287;
Best Local Similarity 43.8%; Pred. No. 0.38;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
Db 63 TOTVFLKKDWDGVKT 78

RESULT 34
US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match          53.6%; Score 52; DB 3; Length 287;
Best Local Similarity 43.8%; Pred. No. 0.38;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
Db 63 TOTVFLKKDWDGVKT 78

RESULT 35
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
```

Mon Oct 6 09:43:09 2003

us-09-765-739a-6.rai

Page 10

APPLICANT: Mahan, Suman M.
TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
FILE REFERENCE: UF-167C1
CURRENT APPLICATION NUMBER: US/08/953,326
EARLIER FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/953,326
EARLIER FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/733,230
EARLIER FILING DATE: 1996-10-17
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 286
TYPE: PRT
ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 52.6%; Score 51; DB 3; Length 286;
Best Local Similarity 50.0%; Pred. NO. 0.56;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
Qy 1 NPTVALYGLKQDWE 14
Db 60 NTVGVFGIEQDWD 73

Search completed: October 6, 2003, 07:53:55
Job time : 11.4776 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 9.67164 Seconds
(without alignments)
178.981 Million cell updates/sec

Title: US-09-765-739A-6

Perfect score: 97

Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_76:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	100.0	278	2 JE0216	28k surface antigen
2	62	63.9	280	2 JE0217	28k surface antigen
3	58	59.8	276	2 JE0218	28k surface antigen
4	56	57.7	284	2 I40882	major antigenic pr
5	51	52.6	286	2 JE0219	28k surface antigen
6	48	49.5	133	2 JE0221	28k surface antigen
7	46.5	47.9	553	2 G90593	subtilisin, serine
8	46	47.4	375	2 AG0606	hypothetical RNA m
9	45.5	46.9	122	2 G84558	hypothetical prote
10	45	46.4	1077	2 A96533	probable zinc meta
11	44	45.4	162	2 F82084	probable 2-demethy
12	44	45.4	328	2 F71838	probable nadh oxid
13	44	45.4	477	2 F87329	hypothetical prote
14	44	45.4	856	2 G71133	probable alpha-man
15	43.5	44.8	122	2 A71950	hypothetical prote
16	43	44.3	132	2 T44301	hypothetical prote
17	43	44.3	298	2 S75205	hypothetical prote
18	43	44.3	389	2 T43979	hypothetical prote
19	43	44.3	412	2 T09313	immediate-early pr
20	42.5	43.8	370	2 T40131	hypothetical prote
21	42	43.3	161	2 AF0013	S-adenosylmethioni
22	42	43.3	178	2 T29345	hypothetical prote
23	42	43.3	190	2 E95093	hydrolase, haloaci
24	42	43.3	190	2 B97961	phosphoglycolate p
25	42	43.3	297	2 AH3649	iron(III)-transpor
26	42	43.3	373	2 F75073	hypothetical prote
27	42	43.3	757	2 C84120	subtilisin-type pr
28	42	43.3	1023	2 A47296	thiazide-sensitive
29	41	42.3	168	2 S18374	flavodoxin - Anaba

flavodoxin - Anaba	30	41	42.3	170	1 S04600
flavodoxin [import	31	41	42.3	170	2 AF2106
alanine racemase [32	41	42.3	368	2 AD0937
conserved hypothe	33	41	42.3	395	2 H72222
phosphoglycerate t	34	41	42.3	439	2 E82426
hypothetical prote	35	41	42.3	452	2 D70768
hypothetical prote	36	41	42.3	480	2 T15839
CTP synthase (EC 6	37	41	42.3	544	2 S25101
regulatory protein	38	41	42.3	918	2 S04255
dna exoribonuclea	39	41	42.3	991	2 S43891
hypothetical prote	40	40.5	41.8	180	2 T16668
hypothetical prote	41	40	41.2	105	2 F83935
hypothetical prote	42	40	41.2	135	2 G83495
antirestriction pr	43	40	41.2	169	2 A42716
flavodoxin [valida	44	40	41.2	170	1 A28670
orotidine-5'-phosp	45	40	41.2	239	2 I39845
ATP-binding protei	46	40	41.2	254	2 F89309
hypothetical prote	47	40	41.2	341	2 S73685
hypothetical prote	48	40	41.2	374	2 H71091
probable starch sy	49	40	41.2	486	2 C96018
probable membrane	50	40	41.2	661	2 S45131
bumetanide sensiti	51	40	41.2	774	2 G71308
hypothetical prote	52	40	41.2	1060	2 T30823
probable RNA helic	53	40	41.2	1122	2 T28130
pre-mRNA splicing	54	40	41.2	1220	2 A56236
hypothetical prote	55	40	41.2	1226	2 T49915
hypothetical prote	56	40	41.2	1587	2 AB2012
carbamoyl-phosphat	57	39.5	40.7	358	2 B64427
hypothetical prote	58	39	40.2	99	2 G89891
hypothetical prote	59	39	40.2	214	2 T27259
modulation protein	60	39	40.2	226	1 JQ0393
hypothetical prote	61	39	40.2	231	2 T32146
probable ABA-repo	62	39	40.2	233	2 G85020
conserved hypotet	63	39	40.2	238	2 T08289
kanamycin kinase (64	39	40.2	263	1 PKSAF
kanamycin kinase (65	39	40.2	264	1 PKSOJF
kanamycin resistanc	66	39	40.2	264	2 I40613
metallopeptidase h	67	39	40.2	264	2 T03174
rRNA N-glycosidase	68	39	40.2	286	1 RLPUGG
hypothetical prote	69	39	40.2	302	2 S75227
rod shape-determin	70	39	40.2	335	2 B82220
glycosyltransferas	71	39	40.2	363	2 D69502
hypothetical prote	72	39	40.2	382	2 AF2994
hypothetical prote	73	39	40.2	382	2 D98289
phosphoglycerate t	74	39	40.2	392	2 A28255
phosphoglycerate t	75	39	40.2	415	2 AG0806
probable membrane	76	39	40.2	461	2 AC0005
gp10 protein - Myc	77	39	40.2	493	2 G72800
hypothetical prote	78	39	40.2	525	2 A72586
probable Acyl-CoA	79	39	40.2	580	2 B70668
probable acid-CoA	80	39	40.2	585	2 B75265
64k capsid assembl	81	39	40.2	590	1 WMBECB
polyprotein - equi	82	39	40.2	643	2 S55610
DNA ligase - Therm	83	39	40.2	688	2 D72418
hypothetical prote	84	39	40.2	691	2 B75622
fused proteinase-c	85	39	40.2	697	2 JC2365
translation elonga	86	39	40.2	732	2 A75219
probable translati	87	39	40.2	735	2 G71203
hypothetical prote	88	39	40.2	772	2 E96977
cysteine proteinas	89	39	40.2	886	2 T10890
TonB-dependent rec	90	39	40.2	888	2 B87270
probable penicilli	91	39	40.2	1009	2 D75399
probable relaxase	92	39	40.2	1102	2 T02782
bumetanide-sensiti	93	39	40.2	1191	2 A53491
pyridoxal phosphat	94	38.5	39.7	349	1 S77070
hypothetical prote	95	38.5	39.7	465	2 E69788
cyclomaltodextrin	96	38.5	39.7	710	2 S63598
probable transmem	97	38.5	39.7	766	2 B75059
hypothetical prote	98	38	39.2	85	2 T27307
hypothetical prote	99	38	39.2	131	2 C72491
hypothetical prote	100	38	39.2	165	2 T47011

ALIGNMENTS

RESULT 1

JE0216
28k surface antigen 3 - Ehrlichia chaffensis
N;Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0216; MUID:98321180; PMID:9647746
A;Accession: JE0216
A;Molecule type: DNA
A;Residues: 1-278 <RED>
A;Cross-references: GB:AF062761

Query Match 100.0%; Score 97; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 3e-08; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0;

QY 1 NPTVALYGLKQDWEGISS 18
| | | | | | | | | | | | | | | | | | | | | |
Db 60 NPTVALYGLKQDWEGISS 77

RESULT 2

JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N;Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0217; MUID:98321180; PMID:9647746
A;Accession: JE0217
A;Molecule type: DNA
A;Residues: 1-280 <RED>
A;Cross-references: GB:AF062761

Query Match 63.9%; Score 62; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.017; Mismatches 4; Indels 2; Gaps 0;
Matches 9; Conservative 4;

QY 1 NPTVALYGLKQDWEG 15
| | | | | | | | | | | | | | | | | | | | | |
Db 60 NPTVALYGLKQDWEG 74

RESULT 3

JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N;Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0218; MUID:98321180; PMID:9647746
A;Accession: JE0218
A;Molecule type: DNA
A;Residues: 1-276 <RED>
A;Cross-references: GB:AF062761

Query Match 59.8%; Score 58; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.078; Mismatches 4; Indels 2; Gaps 0;
Matches 9; Conservative 4;

QY 1 NPTVALYGLKQDWEG 15
| | | | | | | | | | | | | | | | | | | | | |
Db 59 NPTVALYGLKQDWEG 73

RESULT 4

JE0219
major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R;van Vliet, A.H.; Jongschaap, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A;Title: Molecular cloning, sequence analysis, and expression of the gene encoding the
A;Reference number: I40882; MUID:94178956; PMID:8132352
A;Accession: I40882
A;Status: Preliminary; translated from GB/EMBL/DDBB
A;Molecule type: DNA
A;Residues: 1-284 <RES>
A;Cross-references: EMBL:X74250; NID:9454266; PIDN:CAA52309.1; PID:9454267
C:Genetics:
A:Gene: map1

Query Match 57.7%; Score 56; DB 2; Length 284;
Best Local Similarity 50.0%; Pred. No. 0.17; Mismatches 6; Indels 2; Gaps 0;
Matches 8; Conservative 6;

QY 3 TVALYGLKQDWEGISS 18
| | | | | | | | | | | | | | | | | | | | | |
Db 63 TKAFLGLKQDWGVKT 78

RESULT 5

JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N;Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0219; MUID:98321180; PMID:9647746
A;Accession: JE0219
A;Molecule type: DNA
A;Residues: 1-286 <RED>
A;Cross-references: GB:AF062761

Query Match 52.6%; Score 51; DB 2; Length 286;
Best Local Similarity 50.0%; Pred. No. 1.1; Mismatches 5; Indels 2; Gaps 0;
Matches 7; Conservative 5;

QY 1 NPTVALYGLKQDWE 14
| | | | | | | | | | | | | | | | | | | | | |
Db 60 NPTVALYGLKQDWE 73

RESULT 6

JE0221
28k surface antigen 2 - Ehrlichia canis
C:Species: Ehrlichia canis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
C:Accession: JE0221
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0221; MUID:98321180; PMID:9647746
A;Accession: JE0221
A;Molecule type: DNA
A;Residues: 1-133 <RED>
A;Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC26722.1; PID:g3327966

Query Match 49.5%; Score 48; DB 2; Length 133;
Matches 4; Indels 2; Gaps 0;

Best Local Similarity 55.6%; Pred. No. 1.6;
Matches 10; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 3 TVALYGLKQDWEG--ISS 18
| .|.|.|.|.|.|.|.|.|.
DB 62 TTVVYGLKENWAGDAISS 79

RESULT 7
G90593
subtilisin, serine proteinase [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: G90593
R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulmonis
A:Reference number: A99512; MUID:21267165; PMID:11353084
A:Accession: G90593
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-553 <KUR>
A:Cross-references: GB:AL445566; PID:gl4090070; PIDN:CAC13828.1; GSPDB:GN00153
A:Experimental source: strain UAB CTIP
C:Genetics:
A:Gene: MYP0_6550
A:Genetic code: SGC3

Query Match 47.9%; Score 46.5; DB 2; Length 553;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 9; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 NPTVALYGLKQ-DWEGIS 17
| .|.|.|.|.|.|.|.|.
DB 229 NERIKLYGKRFNNNGIS 246

RESULT 8
AG0606
hypothetical RNA methyltransferase (EC 2.1.1.1-) [imported] - Salmonella enterica subsp.
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0606
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0606
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05321.1; PID:gl6502085; GSPDB:GN00176
C:Genetics:
A:Gene: ybfJ
C:Superfamily: hypothetical protein HI0333
C:Keywords: methyltransferase

Query Match 47.4%; Score 46; DB 2; Length 375;
Best Local Similarity 58.8%; Pred. No. 10;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 NPTVA--LYLGLKQDWEG 15
|.|.|.|.|.|.|.|.
DB 214 NPTVASRLYATARDWVG 230

RESULT 9
G64558
hypothetical protein HP0311 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: G64558
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee,
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watther,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: G64558
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-122 <TON>
A:Cross-references: GB:AE000549; GB:AE000511; NID:g2313403; PIDN:AAD07384.1; PID:g2313403

Query Match 46.9%; Score 45.5; DB 2; Length 122;
Best Local Similarity 56.2%; Pred. No. 3.8;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 NPTVALYGLK-QDWEG 15
| .|.|.|.|.|.|.
DB 55 NQILAFYGLKINDWQG 70

RESULT 10
A96533
probable zinc metalloproteinase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A96533
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzita
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A96533
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1077 <SPO>
A:Cross-references: GB:AE005173; NID:gl0120424; PIDN:AAG13049.1; GSPDB:GN00141
C:Genetics:
A:Gene: Fl4J22.13
A:Map position: 1

Query Match 46.4%; Score 45; DB 2; Length 1077;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 QDWEGISS 18
|.|.|.|.|.|.
DB 817 QDWEGISS 824

RESULT 11
F82084
probable 2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.1-) VC2366 [similarity]
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82084
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: F82084

Mon Oct 6 09:43:10 2003

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-162 <HEI>
A:Cross-references: GB:AE004307; GB:AE003852; NID:g9656934; PIDN:AAF95509.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype E1 Tor
C:Genetics:
A:Gene: VC2366
A:Map position: 1
C:Keywords: methyltransferase

Query Match 45.4%; Score 44; DB 2; Length 162;
Best Local Similarity 46.2%; Pred. No. 9;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 VALYGLKQDWEG 16
Db 79 LAILAINKDWEV 91

RESULT 12
G71838
probable nadh oxidoreductase I - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: G71838
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: G71838
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-328 <ARN>
A:Cross-references: GB:AE001545; GB:AE001439; NID:g4155776; PIDN:AAD06772.1; PID:g415579
A:Experimental source: strain J99
C:Genetics:
A:Gene: nuof

Query Match 45.4%; Score 44; DB 2; Length 328;
Best Local Similarity 50.0%; Pred. No. 19;
Matches 12; Conservative 1; Mismatches 3; Indels 8; Gaps 2;

QY 1 NPTVALYGLK-QDW-----EGI 16
Db 203 NPTVAFYDSKDQEWLLETFKEGI 226

RESULT 13
F87329
hypothetical protein CC0649 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: F87329
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.;
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87329
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-477 <STO>
A:Cross-references: GB:AE005673; NID:g13421864; PIDN:AAK22634.1; GSPDB:GN00148
A:Gene: CC0649

Query Match 45.4%; Score 44; DB 2; Length 477;
Best Local Similarity 57.1%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 2 PTVALYGLKQDWEG 15
Db 254 PTWGLYNSKWDWRG 267

RESULT 14
G71133
probable alpha-mannosidase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: G71133
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, I.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: G71133
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-856 <KAW>
A:Cross-references: GB:AP000003; NID:g3236130; PIDN:BAA29929.1; PID:d1030872; PID:g32
A:Experimental source: Strain OF3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH0835

Query Match 45.4%; Score 44; DB 2; Length 856;
Best Local Similarity 63.6%; Pred. No. 52;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 LYGLKQDWEGI 16
Db 309 LYGIETPWEGI 319

RESULT 15
A71950
hypothetical protein jhp0296 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71950
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <ARN>
A:Cross-references: GB:AE001466; GB:AE001439; NID:g4154813; PIDN:AAD05873.1; PID:g41
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0296

Query Match 44.8%; Score 43.5; DB 2; Length 122;
Best Local Similarity 56.2%; Pred. No. 8.1;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 NPTVALYGLK-ODWEG 15
Db 55 NQILAFYGLKIGDWOG 70

RESULT 16
T44301
hypothetical protein BH0656 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 15-Jun-2001
C:Accession: T44301; H83731
R:Rakani, H.; Nakasone, K.; Ogasawara, N.; Hirama, C.; Nakamura, Y.; Masui, N.; Fuji
Extremophiles 3, 29-34, 1999

```

A:Title: Sequencing of three lambda clones from the genome of alkaliphilic Bacillus sp.
A:Reference number: Z22745; MUID:99184646; PMID:10086842
A:Accession: T44301
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-132 <TAK>
A:Cross-references: EMBL:AB011836; NID:g4512345; PIDN:BA075320.1; PID:g4512355
A:Experimental source: strain C-125
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: AB3650; MUID:20512582; PMID:11058132
A:Accession: H83731
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-132 <STO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04375.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
C:Gene: BH0556
C:Superfamily: hypothetical protein b2531

Query Match 44.3%; Score 43; DB 2; Length 132;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 YGLKODWEGIS 17
I I I I I I I I
Db 65 YLKRDEWEIS 75

RESULT 17
S75205
hypothetical protein sir2052 - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75205
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S75205
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-298 <KAN>
A:Cross-references: EMBL:D90903; GB:AB001339; NID:g1652127; PIDN:BA017119.1; PID:g165219
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synecocystis hypothetical protein sir2052

Query Match 44.3%; Score 43; DB 2; Length 298;
Best Local Similarity 42.1%; Pred. No. 25;
Matches 8; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 2 PTVALYG--LKODWEGISS 18
I I I I I I I I I
Db 278 PTVIAYGKRISRSWSGVQS 296

RESULT 18
T43979
hypothetical protein U19 [imported] - human herpesvirus 6
C:Species: human herpesvirus 6
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000
C:Accession: T43979; T44165
R:Isegawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; Kawa
J. Virol. 73, 8053-8063, 1999
A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A and
A:Reference number: Z22732; MUID:99412319; PMID:10482554
A:Accession: T43979
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-389 <ISE>
A:Cross-references: EMBL:AB021506; NID:g4959577; PIDN:BA078240.1; PID:g49596007
A:Experimental source: strain HST; pop. variant B
R:Dominguez, G.; Dambaugh, F.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.
J. Virol. 73, 8040-8052, 1999
A:Title: Human herpesvirus 6B genome sequence: coding content and comparison with human
A:Reference number: Z22734; MUID:99412318; PMID:10482553
A:Accession: T44165
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-389 <DOM>
A:Cross-references: EMBL:AF157706; PIDN:AA049631.1
A:Experimental source: strain Z29; variant B
C:Genetics:
C:Gene: U19
C:Superfamily: human herpesvirus 6 hypothetical protein U19

Query Match 44.3%; Score 43; DB 2; Length 389;
Best Local Similarity 33.3%; Pred. No. 33;
Matches 5; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

QY 4 VALYGLKODWEGISS 18
:::|: ::||| I:
Db 79 MSMFGVTEWEGASA 93

RESULT 19
T09313
Immediate-early protein 4 - human herpesvirus 6 (strain U1102)
C:Species: human herpesvirus 6
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 28-Jul-2000
C:Accession: T09313
R:Nicholas, J.; Martin, M.
J. Virol. 68, 597-610, 1994
A:Title: Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of
A:Reference number: Z16644; MUID:94118404; PMID:8289364
A:Accession: T09313
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-412 <NIC>
A:Cross-references: EMBL:L25528; NID:g451932; PIDN:AAA16726.1; PID:g451944
C:Genetics:
C:Gene: EJLF4
C:Superfamily: human herpesvirus 6 hypothetical protein U19

Query Match 44.3%; Score 43; DB 2; Length 412;
Best Local Similarity 33.3%; Pred. No. 35;
Matches 5; Conservative 8; Mismatches 2; Indels 0; Gaps 0;

QY 4 VALYGLKODWEGISS 18
:::|: ::||| I:
Db 102 MSMFGVTEWEGASA 116

RESULT 20
T40131
hypothetical protein SPBC2F12.10 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40131
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, June 1997
A:Reference number: Z21907
A:Accession: T40131
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-370 <WOO>
A:Cross-references: EMBL:Z97211; PIDN:CAB10157.1; GSPDB:GN00067; SPDB:SPBC2F12.10
A:Experimental source: strain 972h-; cosmid c2F12
C:Genetics:
C:Gene: SPDB:SPBC2F12.10
A:Map position: 2

Query Match 43.8%; Score 42.5; DB 2; Length 370;
 Best Local Similarity 71.4%; Pred. No. 38;
 Matches 10; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
 QY 1 NPTVALYGLKQDWE 14
 Db 151 NP-VFLYMLKQDWE 163
 RESULT 21
 AF0013
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AF0013
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AF0013
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-161 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC88968.1; PID:gl15978212; GSPDB:GN00175
 C:Genetics:
 A:Gene: meng
 C:Keywords: transferase

Query Match 43.3%; Score 42; DB 2; Length 161;
 Best Local Similarity 50.0%; Pred. No. 19;
 Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWE 16
 Db 75 NAEALALKNEWEG 90

RESULT 22
 T29345
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29345
 R:Du, Z.; Gattung, S. submitted to the EMBL Data Library, April 1996
 A:Description: The sequence of C. elegans cosmid F21C10.
 A:Reference number: Z20610.
 A:Accession: T29345
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-178 <DUZ>
 A:Cross-references: EMBL:U53364; PIDN:AAA97978.1; GSPDB:GN00023; CESP:F21C10.2
 A:Experimental source: strain Bristol N2; clone F21C10
 C:Genetics:
 A:Gene: CESP:F21C10.2
 A:Map position: 5
 A:Introns: 34/2; 70/3; 107/2

Query Match 43.3%; Score 42; DB 2; Length 178;
 Best Local Similarity 58.3%; Pred. No. 21;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 VALYGLKQDWE 15
 Db 36 ISLYGQKYWEG 47

RESULT 23
 E95093
 C:Species: Streptococcus pneumoniae (family [imported]) - Streptococcus pneumoniae (stra
 hydrolase, haloacid dehalogenase-like family [imported]) - Streptococcus pneumoniae (stra

C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
 C:Accession: E95093
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Henson, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf, T.; Hickey, E.K.; Holt, I.E. Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris, A.; Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: E95093
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-190 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK74942.1; PID:gl14972282; GSPDB:GN00164; TIGR: A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0805

Query Match 43.3%; Score 42; DB 2; Length 190;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWE 16
 Db 31 TLALYGLTQDHSV 44

RESULT 24
 B97961
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: B97961
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.Y.; P.; Sun, P.M.; Winkler, M.E. J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.A.; Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: B97961
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-190 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK99518.1; PID:gl15458305; GSPDB:GN00174
 C:Genetics:
 A:Gene: gph
 C:Keywords: phosphoric monoester hydrolase

Query Match 43.3%; Score 42; DB 2; Length 190;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWE 16
 Db 31 TLALYGLTQDHSV 44

RESULT 25
 AH3649
 C:Species: Streptococcus pneumoniae (family [imported]) - Streptococcus pneumoniae (stra
 iron(III)-transport system permease protein sfub [imported]) - Brucella melitensis (s
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AH3649
 R:Deiveccchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Los, T.; Ivancv
 Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A:Reference number: AD3252; MUID:11756688
 A:Accession: AH3649
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-297 <KUR>
 A:Cross-references: GB:AE008918; PIDN:AAL54363.1; PID:g17985346; GSPDB:GN00191
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI1121
 A:Map position: II

Query Match 43.3%; Score 42; DB 2; Length 297;
 Best Local Similarity 56.2%; Pred. No. 36;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWEGISS 18
 ||| |||:|:|:| ||
 Db 182 TPALDGLKKEWREASS 197

RESULT 26
 G75073
 hypothetical protein PAB1719 - Pyrococcus abyssi (strain Orsay)
 C:Species: Pyrococcus abyssi
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 28-Jul-2000
 C:Accession: G75073
 R:Anonymous, Genoscope
 submitted to the EMBL Data Library, July 1999
 A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure
 A:Reference number: A75001
 A:Accession: G75073
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-373 <KAW>
 A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB49892.1; PID:g545840
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: PAB1719
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0997

Query Match 43.3%; Score 42; DB 2; Length 373;
 Best Local Similarity 58.3%; Pred. No. 46;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWE 14
 ||:|:| | |:
 Db 87 TVSVYGRKIDWK 98

RESULT 27
 CB4120
 subtilisin-type proteinase (EC 3.4.21.-) BH3763 precursor [similarity] - Bacillus halodurans
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: CB4120
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: CB4120
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-757 <STO>
 A:Cross-references: GB:AF001519; GB:BA000004; NID:g10176109; PIDN:BA07482.1; GSPDB:GN00
 A:Experimental source: strain C-125.
 C:Genetics:
 A:Gene: BH3763
 C:Superfamily: microbial serine proteinase vpr; subtilisin homology
 C:Keywords: hydrolase; serine proteinase
 F:1-20/Domain: signal sequence #status predicted <SIG>

Query Match 43.3%; Score 42; DB 2; Length 757;
 Best Local Similarity 41.7%; Pred. No. 97;
 Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 7 YGLKQDWEGISS 18
 || :|:|:|:|:

Db 336 YGTEKEWEGVDA 347

RESULT 28
 AA7296
 thiazide-sensitive electroneutral sodium-chloride cotransporter - winter flounder
 C:Species: pseudopleuronectes americanus (winter flounder)
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Mar-2001
 C:Accession: AA7296
 R:Gamba, G.; Saltberg, S.N.; Lombardi, M.; Miyanoshta, A.; Lytton, J.; Hediger, M.A.
 Proc. Natl. Acad. Sci. U.S.A. 90, 2749-2753, 1993
 A:Title: Primary structure and functional expression of a cDNA encoding the thiazide-
 A:Reference number: A47296; MUID:93219361; PMID:8464884
 A:Accession: AA7296
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-1023 <GAM>
 A:Experimental source: urinary bladder
 A:Note: sequence extracted from NCBI backbone (NCBIN:128719, NCBIP:128720)
 C:Superfamily: rat bumetanide-sensitive Na+/K+/Cl-cotransport protein

Query Match 43.3%; Score 42; DB 2; Length 1023;
 Best Local Similarity 58.3%; Pred. No. 1.3e+02;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PTVALLGLKQDW 13
 | | | | |:
 Db 727 PNVLLMGFKQDW 738

RESULT 29
 SI8374
 flavodoxin - Anabaena sp. (PCC 7119) (fragment)
 C:Species: Anabaena sp.
 A:Variety: PCC 7119
 C:Date: 22-Nov-1993 #sequence_revision 08-Nov-1996 #text_change 25-Apr-1997
 C:Accession: SI8374; S20298
 R:Fillat, M.F.; Borrias, W.E.; Weisbeek, P.J.
 Biochem. J. 280, 187-191, 1991
 A:Title: Isolation and overexpression in Escherichia coli of the flavodoxin gene from
 A:Reference number: SI8374; MUID:92074973; PMID:1720613
 A:Accession: SI8374
 A:Molecule type: DNA
 A:Residues: 1-168 <FIL>
 A:Experimental source: Anabaena sp. PCC 7119
 R:Medina, M.; Peleato, M.L.; Mendez, E.; Gomez-Moreno, C.
 Eur. J. Biochem. 203, 373-379, 1992
 A:Title: Identification of specific carboxyl groups on Anabaena PCC 7119 flavodoxin v

A:Accession: S20298
 A:Molecule type: protein
 A:Residues: 115-128;137-144 <MED>
 A:Experimental source: Anabaena sp. PCC 7119
 C:Superfamily: flavodoxin; flavodoxin homology
 C:Keywords: electron transfer; flavoprotein; FMN
 F:5-163/Domain: flavodoxin homology <FLX>

Query Match 42.3%; Score 41; DB 2; Length 168;
 Best Local Similarity 47.1%; Pred. No. 29;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 PTVALLGLKQDWEGISS 18
 | | : | : | | |:
 Db 54 PTWNIGELQSDWEGLYS 70

RESULT 30
 S04600
 flavodoxin - Anabaena variabilis
 C:Species: Anabaena variabilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S04600; S05277
 R:Leonhardt, K.G.; Straus, N.A.

Search completed: October 6, 2003, 07:49:42
Job time : 12.6716 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:03 ; Search time 33.6045 Seconds
(without alignments)
89.744 Million cell updates/sec

Title: US-09-765-739A-7

Perfect score: 103

Sequence: 1 NNTTGVFLKQWDGDTIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	103	100.0	19	ABG30749	Ehrlichia chaffeen
2	103	100.0	280	AAU06948	E. chafeensis OMP-
3	103	100.0	280	ABG77940	Ehrlichia chaffeen
4	103	100.0	280	AAU96110	Ehrlichia chaffeen
5	97	94.2	280	AAW51094	Ehrlichia chaffeen
6	97	94.2	280	AAAB36188	Ehrlichia chaffeen
7	97	94.2	280	AAU04198	Variable surface a
8	97	94.2	280	AAU73417	Ehrlichia chaffeen
9	96	93.2	20	ABG30744	Ehrlichia canis pe

10	96	93.2	288	20	AAU06959	E. canis p30 prote
11	96	93.2	288	23	ABG77950	Ehrlichia canis ou
12	88	85.4	20	23	ABG30745	Ehrlichia chaffeen
13	88	85.4	256	20	AAU06942	E. chafeensis p28
14	88	85.4	256	23	ABG77966	Protein encoded by
15	88	85.4	276	19	AAW51095	Ehrlichia chaffeen
16	88	85.4	276	21	AAAB36189	Ehrlichia chaffeen
17	88	85.4	276	22	AAU04199	Variable surface a
18	88	85.4	280	19	AAW51089	Ehrlichia chaffeen
19	88	85.4	280	21	AAAB36183	Ehrlichia chaffeen
20	88	85.4	280	22	AAU04193	Major antigenic pr
21	88	85.4	281	20	AAU06943	E. chafeensis OMP-
22	88	85.4	281	23	ABG77935	Ehrlichia chaffeen
23	88	85.4	281	23	AAU96105	Ehrlichia chafeens
24	88	85.4	281	23	AAU73418	Ehrlichia chaffeen
25	80	77.7	280	20	AAU06962	E. canis p30-2 pro
26	80	77.7	280	21	AAU71479	Ehrlichia canis im
27	80	77.7	280	23	ABG77953	Ehrlichia canis ou
28	80	77.7	280	23	AAU96102	Ehrlichia canis p2
29	74	71.8	19	23	ABG30747	Ehrlichia chaffeen
30	74	71.8	286	19	AAW51092	Ehrlichia chaffeen
31	74	71.8	286	20	AAU06946	E. chafeensis OMP-
32	74	71.8	286	21	AAAB36186	Ehrlichia chaffeen
33	74	71.8	286	22	AAU04196	Variable surface a
34	74	71.8	286	23	ABG77938	Ehrlichia chaffeen
35	74	71.8	286	23	AAU96108	Ehrlichia chafeens
36	74	71.8	286	23	AAU73415	Ehrlichia chaffeen
37	73	70.9	20	23	ABG30743	Ehrlichia canis pe
38	73	70.9	278	21	AAU71477	Ehrlichia canis im
39	73	70.9	278	23	AAU96100	Ehrlichia canis p2
40	73	70.9	307	20	AAU06961	E. canis p30-1 pro
41	73	70.9	307	23	ABG77952	Ehrlichia canis ou
42	72	69.9	276	20	AAU06964	Ehrlichia canis ou
43	72	69.9	276	23	ABG77955	Ehrlichia canis p2
44	72	69.9	276	23	AAU96117	Ehrlichia chaffeen
45	60	58.3	278	19	AAW51093	E. chafeensis OMP-
46	60	58.3	278	20	AAU06947	Ehrlichia chaffeen
47	60	58.3	278	21	AAAB36187	Ehrlichia chaffeen
48	60	58.3	278	22	AAU04197	Variable surface a
49	60	58.3	278	23	AAE31090	Ehrlichia ruminant
50	60	58.3	278	23	ABG77939	Ehrlichia chaffeen
51	60	58.3	278	23	AAU96109	Ehrlichia chafeens
52	60	58.3	278	23	AAU73416	Ehrlichia chaffeen
53	60	58.3	280	20	AAU06945	E. chafeensis OMP-
54	60	58.3	280	23	ABG77937	Ehrlichia chaffeen
55	60	58.3	280	23	AAU96107	Ehrlichia chafeens
56	60	58.3	280	23	AAU73414	Ehrlichia chaffeen
57	59.5	57.8	287	19	AAW51088	Cowdria ruminantiu
58	59.5	57.8	287	21	AAAB36182	Cowdria ruminantiu
59	59.5	57.8	287	22	AAU04192	Major antigenic pr
60	59	57.3	18	23	ABG30748	Ehrlichia chaffeen
61	59	57.3	284	23	AAU96111	Cowdria ruminantiu
62	58	56.3	18	23	ABG30746	Ehrlichia chaffeen
63	55	53.4	132	22	AAU04201	Variable surface a
64	55	53.4	133	19	AAW51097	Ehrlichia canis VS
65	55	53.4	133	21	AAAB36191	Ehrlichia canis pa
66	55	53.4	133	21	AAU71480	Ehrlichia canis im
67	55	53.4	133	23	AAU96103	Ehrlichia canis p2
68	55	53.4	283	21	AAU71478	Ehrlichia canis im
69	55	53.4	283	23	AAU96101	Ehrlichia canis p2
70	52	50.5	240	21	AAG20226	Arabidopsis thalia
71	52	50.5	364	21	AAG47084	Arabidopsis thalia
72	52	50.5	365	21	AAG47111	Arabidopsis thalia
73	52	50.5	374	21	AAG47083	Arabidopsis thalia
74	52	50.5	375	21	AAG47110	Arabidopsis thalia
75	52	50.5	442	21	ABG47082	Arabidopsis thalia
76	52	50.5	443	21	ABG47109	Arabidopsis thalia
77	47	45.6	65	22	AAU63107	Propionibacterium
78	44	42.7	308	22	ABBS8819	Propionibacterium
79	44	42.7	1648	23	ABBS4925	Lactococcus lactis
80	43	41.7	806	22	AAAG92136	C glutamicum prote
81	43	41.7	904	20	AAU21976	Senescence-associa
82	43	41.7	1429	22	ABBS9606	Drosophila melanog

83	42	40.8	107	23	AAOL17077	Human acyl CoA syn
84	42	40.8	459	23	AB965331	Bifidobacterium lo
85	41.5	40.3	467	23	AB949886	Listeria monocytog
86	41	39.8	79	22	ABG13616	Novel human diagn
87	41	39.8	135	25	AA955866	Melanogenic inhibi
88	41	39.8	135	22	AAU08674	Human keratinocyte
89	41	39.8	135	23	ABP65127	Hypoxia-repressed
90	41	39.8	135	23	ABG61772	Novel fatty acid-b
91	41	39.8	155	22	AAU00890	Human cancer relat
92	41	39.8	158	22	ABG27577	Novel human diagn
93	41	39.8	198	21	AAO73139	Arabidopsis thalia
94	41	39.8	198	21	AAO38298	Arabidopsis thalia
95	41	39.8	210	22	ABG22152	Novel human diagn
96	41	39.8	244	21	AAO07318	Arabidopsis thalia
97	41	39.8	244	21	AAO38297	Arabidopsis thalia
98	41	39.8	375	23	AB949335	Listeria monocytog
99	41	39.8	424	21	AAO18711	Arabidopsis thalia
100	41	39.8	424	21	AAO48446	Arabidopsis thalia

ALIGNMENTS

RESULT 1
ABG30749
ID ABG30749 standard; peptide: 19 AA.

21-OCT-2002 (first entry)

Ehrlichia chaffeensis peptide fragment #5.

KW Antibody detection; monoclonal antibody; polyclonal antibody.

OS Ehrlichia chaffeensis.

PN WO200257794-A2.

25-JUL-2002.

16-JAN-2002; 2002WO-US01395.

PR 18-JAN-2001; 2001US-0765739.

PA (IDEX-) IDEXX LAB INC.

PI
AA
Lawton R, O'Connor TP, Bartol BA, Machenry PS;

WPI; 2002-599730/64.

New composition of matter comprising a polypeptide, useful in detecting the presence of antibodies to *Ehrlichia canis* or *chaffeensis*, or in detecting or quantifying the presence of *Ehrlichia* infection in mammals

XX
PS
Claim 1: Page 5: 29pp: English.

The invention relates to a composition of matter comprising a polypeptide isolated from *Ehrlichia* species. The composition can be used for detecting the presence of antibodies to *Ehrlichia*, comprising contacting one or more polypeptides with a test sample suspected of comprising antibodies to *Ehrlichia*, under conditions that allow polypeptide/antibody complexes to form and detecting the complexes, where the detection of polypeptide/antibody complexes is an indication that antibodies to *Ehrlichia* are present in the test sample. The composition is useful for detecting or quantifying the presence of *E. canis* or *E. chaffeensis* infection in mammals. The polypeptides can be used to develop monoclonal and/or polyclonal antibodies that can be employed in assay systems and in the generation of chimeric antibodies for therapeutic use or other similar applications. This sequence represents an *E. chaffeensis* peptide fragment used in the composition of the invention.

SQ Sequence 19 AA;

RESULT 2

AA100540
ID AAY06948 standard; Protein

AA
AC
AA06948:

DT 05-JUL-1999 (first entry)

DE *E. chafeensis* OMP-1F

Outer membran

XX
XX
NOT 333333

XX
XII

XX
XX
NN 02560M
-FI.

FD
XX
ZJ-MAR-1999.

XX
11 OCT 1966
XX

FR 19-SEP-1997; 9703-
XX

XX
XX
OTHO ATNO (ETHO)

PI OMASHT N, RIKINISA I;
XX

DR WPI; 1999-254290/21.
DR N-PSDB: AAX34748

Novel outer membrane proteins from *Ehrlichia chaffeensis* and *Ehrlichia canis*

Claim 16: Fig 8B; 55pp; English.

The invention provides isolated outer membrane proteins (OMP) from *Ehrlichia chaffeensis* and *E. canis*. The *E. chaffeensis* proteins form part of the OMP family and consist of proteins OMP-1, -(B to Z) shown in AAY06943-958. The *E. canis* proteins form part of the p30 family and consist of proteins shown in AAY06959-970. The proteins and genes are used to detect *E. chaffeensis* in patients and *E. canis* in dogs.

Sequence	280 AA;
SQ	

RESULT 3

ABG 77940	ABG 77940 standard; Protein
ID	ABG77940 standard; Protein

AA
AC
ABG77940;

KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
 XX Ehrlichia chaffeensis.
 OS
 XX US2002120115-A1.
 PN
 XX 29-AUG-2002.
 PD
 XX 28-JAN-2002; 2002US-0059964.
 PF
 XX 19-MAY-1999; 99US-0314701.
 PR
 XX (RIKI/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.
 XX
 PI Rikihisa Y, Ohashi N;
 XX
 XX WPI; 2002-618954/66.
 DR N-PSDB; ABS63281.
 XX
 PT Isolated polynucleotide encoding an outer membrane protein of E.canis
 PT or E.chaffeensis used in the diagnosis of infection -
 XX
 PS Disclosure; Fig 8B; 49pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.
 XX
 XX Sequence 280 AA;
 SQ
 Query Match 100.0%; Score 103; DB 23; Length 280;
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NTTTGVFGLKQDWGSGTIS 19
 |||||||
 DB 60 NTTTGVFGLKQDWGSGTIS 78
 RESULT 4
 AAU96110
 ID AAU96110 standard; Protein; 280 AA.
 XX
 AC AAU96110;
 XX
 DT 02-JUL-2002 (first entry)
 DE Ehrlichia chaffeensis OMP-1F.
 XX
 KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
 KW antibacterial.
 XX
 XX Ehrlichia chaffeensis.
 OS
 XX WO200222782-A2.
 PN
 XX 21-MAR-2002.
 PD
 XX 12-SEP-2001; 2001WO-US28759.
 PF
 XX 12-SEP-2000; 2000US-0660587.
 PR
 XX (RERE-) RES DEV FOUND.
 PA
 XX

PI Walker DH, Yu X, McBride JW;
 XX WPI; 2002-351882/38.
 DR
 XX New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
 XX
 PS Example 3; Figure 3; 106pp; English.
 XX
 CC The invention relates to a recombinant homologous 28 kDa immunodominant
 CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX
 XX Sequence 280 AA;
 SQ
 Query Match 100.0%; Score 103; DB 23; Length 280;
 Best Local Similarity 100.0%; Pred. No. 1.4e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NTTTGVFGLKQDWGSGTIS 19
 |||||||
 DB 60 NTTTGVFGLKQDWGSGTIS 78
 RESULT 5
 AAWS1094
 ID AAWS1094 standard; Protein; 280 AA.
 XX
 AC AAWS1094;
 XX
 DT 14-SEP-1998 (first entry)
 DE Ehrlichia chaffeensis VSA4 protein.
 XX
 KW MAPI homologue; variable surface antigen; VSA4; rickettsia;
 KW DNA vaccine.
 XX
 OS Ehrlichia chaffeensis.
 FH Key. Location/Qualifiers
 FT Peptide 1..25
 FT /note= "putative signal peptide"
 XX
 PN WO9816554-A1.
 XX
 PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 DR WPI; 1998-251232/22.
 DR N-PSDB; AAU07179.
 XX
 CC Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 PS Claim 3; Fig 2B; 39pp; English.
 XX
 CC This is the full-length variable surface antigen VSA4 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF4) of a genomic locus (see AAU07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the

CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
CC This genomic locus included 5 ORFs encoding similar, but
CC non-identical proteins (see AAW51091-95). A claimed composition
CC comprises a nucleic acid (see AAW07176-82) encoding a polypeptide
CC (see AAW51088-99) that elicits a protective immune response against a
CC rickettsial pathogen. The nucleic acid is used, in human or
CC veterinary medicine, in vaccines to protect against Rickettsia,
CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
CC polypeptides can be used diagnostically to detect antibodies
CC associated with Ehrlichia infection (claimed).
XX Sequence 280 AA;
SQ
Query Match 94.2%; Score 97; DB 19; Length 280;
Best Local Similarity 94.7%; Pred. No. 1.3e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQDWGSGTIS 19
DB 60 NTTGVFGLKQDWGSGTIS 78
RESULT 6
AAB36188
ID AAB36188 standard; Protein: 280 AA.
XX
AC AAB36188;
DT 02-MAR-2001 (first entry)
XX
DE Ehrlichia chaffeensis partial VSA4.
XX
KW Ehrlichia chaffeensis; VSA4: variable surface antigen 4; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworfl; 18hworfl;
KW 3gdorff3.
XX
OS Ehrlichia chaffeensis.
XX
PI WO200065063-A2.
PN
XX
PD 02-NOV-2000.
XX
PF 21-APR-2000; 2000WO-US10886.
XX
PR 22-APR-1999; 99US-0130725.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbl BH, Whitmire MW, Alleman AR;
XX
DR WPI: 2000-679675/66.
DR N-PSDB; AAC68705.
XX
XX New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
PT
XX
PS Claim 3; Page 45-46; 63pp; English.
XX
XX The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
CC used in a vaccine to protect animals or humans against rickettsial
CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworfl, 18hworfl
CC and 3gdorff3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be

CC used to detect the presence of rickettsial nucleic acids.
XX
SQ Sequence 280 AA;
Query Match 94.2%; Score 97; DB 21; Length 280;
Best Local Similarity 94.7%; Pred. No. 1.3e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQDWGSGTIS 19
DB 60 NTTGVFGLKQDWGSGTIS 78
RESULT 7
AAU04198
ID AAU04198 standard; Protein: 280 AA.
XX
AC AAU04198;
DT 23-OCT-2001 (first entry)
XX
DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
XX
KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic; variable surface antigen; VSA.
XX
OS Ehrlichia chaffeensis.
XX
PN US6251872-B1.
XX
PD 26-JUN-2001.
XX
PF 17-OCT-1997; 97US-0953326.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
DR WPI: 2001-424487/45.
DR N-PSDB; AAS07578.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures
XX
PS Example 3; Fig 2A-2B; 30pp; English.
XX
XX The sequence represents the amino acid sequence of variable surface
CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.
XX
SQ Sequence 280 AA;
Query Match 94.2%; Score 97; DB 22; Length 280;
Best Local Similarity 94.7%; Pred. No. 1.3e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQDWGSGTIS 19
DB 60 NTTGVFGLKQDWGSGTIS 78

```
Db      60 NTTIGVFLKQDWGSGTIS 78

RESULT 8
AAU73417
ID AAU73417 standard; Protein; 280 AA.
XX
AC AAU73417;
XX
DT 12-MAR-2002 (first entry)
XX
DE Ehrlichia chaffeensis outer membrane protein P28-18.
XX
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX
OS Ehrlichia chaffeensis.
XX
PN WO200183699-A2.
XX
PD 08-NOV-2001.
XX
PF 01-MAY-2001; 2001WO-US13997.
XX
PR 01-MAY-2000; 2000US-201035P.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X;
XX
WPI; 2002-066527/09.
XX

Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
P28 useful as a vaccine against Ehrlichia chaffeensis
PS Disclosure; Figure 2; 97pp; English.
XX

The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
SQ Sequence 280 AA;

Query Match 94.2%; Score 97; DB 23; Length 280;
Best Local Similarity 94.7%; Pred. No. 1.3e-07;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSGTIS 19
   ||| ||||| ||||| |||||
Db 60 NTTIGVFLKQDWGSGTIS 78

RESULT 9
ABG30744
ID ABG30744 standard; Peptide; 20 AA.
XX
AC ABG30744;
XX
DT 21-OCT-2002 (first entry)
XX
DE Ehrlichia canis peptide fragment #2.
XX
KW Antibody detection; monoclonal antibody; polyclonal antibody.
XX
OS Ehrlichia canis.
XX
PN WO200257794-A2.
XX
PD 25-JUL-2002.
XX
PF 16-JAN-2002; 2002WO-US01395.

18-JAN-2001; 2001US-0765739.
XX
PR (IDEX-) IDEXX LAB INC.
XX
PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX
WPI; 2002-599730/64.
XX

New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
PT
Claim 1; Page 5; 29pp; English.
XX

The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. canis peptide
CC fragment used in the composition of the invention.
XX
SQ Sequence 20 AA;

Query Match 93.2%; Score 96; DB 23; Length 20;
Best Local Similarity 94.4%; Pred. No. 9.3e-09;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSGTI 18
   ||||| ||||| ||||| ||
Db 1 NTTTGVFLKQDWGSGTI 18

RESULT 10
AAU06959
ID AAU06959 standard; Protein; 288 AA.
XX
AC AAU06959;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis P30 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia canis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS ) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
WPI; 1999-254290/21.
XX
DR N-PSDB; AAX34759.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
```

```

PT  Ehrlichia canis
PS  Disclosure; Fig 19B; 55pp; English.
XX
CC  The invention provides isolated outer membrane proteins (OMP) from
CC  Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC  of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC  in AAY06943-958. The E. canis proteins form part of the P30 family and
CC  consist of proteins shown in AAY06959-970. The proteins and genes are
CC  used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ  Sequence 288 AA;

Query Match          93.2%; Score 96; DB 20; Length 288;
Best Local Similarity 94.4%; Pred. No. 2e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY  1 NNTTGVFGLKQDWDGSTI 18
DB  |||||
60 NNTTGVFGLKQDWDGATI 77

RESULT 11
ABG77950
XX  ABCG77950 standard; Protein; 288 AA.
AC  ABG77950;
DT  15-NOV-2002 (first entry)
DE  Ehrlichia canis outer membrane protein (P30F) #1.
KW  Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
OS  Ehrlichia canis.
XX  US2002120115-A1.
PN  29-AUG-2002.
XX
XX  28-JAN-2002; 2002US-0059964.
XX
XX  19-MAY-1999; 99US-0314701.
XX  (RIKI/) RIKIHISA Y.
XX  (OHAS/) OHASHI N.
XX
XX  Rikihisa Y, Ohashi N;
XX  WPI; 2002-618954/66.
XX  N-PSDB; ABS63291.
XX
XX  Isolated polynucleotide encoding an outer membrane protein of E. canis
XX  or E. chaffeensis used in the diagnosis of infection -
XX
XX  Claim 10; Fig 19B; 49pp; English.
XX
XX  The invention relates to an isolated polynucleotide encoding an outer
XX  membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX  in the diagnosis of infection. An infection such as human ehrlichiosis or
XX  canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX  patient, providing a polypeptide or mixture of polypeptides, contacting
XX  the sample with the polypeptide and assaying for the formation of a
XX  complex between antibodies in the serum sample and the polypeptide, where
XX  formation of a complex is indicative of infection with E. chaffeensis.
XX  This sequence represents an Ehrlichia outer membrane protein of the
XX  invention.
XX
SQ  Sequence 288 AA;

Query Match          93.2%; Score 96; DB 23; Length 288;
Best Local Similarity 94.4%; Pred. No. 2e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY  1 NNTTGVFGLKQDWDGSTI 18
DB  |||||
60 NNTTGVFGLKQDWDGATI 77

RESULT 12
ABG30745
XX  ABG30745 standard; Peptide; 20 AA.
XX
AC  ABG30745;
XX
XX  21-OCT-2002 (first entry)
XX
XX  Ehrlichia chaffeensis peptide fragment #1.
XX
XX  Antibody detection; monoclonal antibody; polyclonal antibody.
XX
XX  Ehrlichia chaffeensis.
XX
XX  WO200257794-A2.
XX
XX  25-JUL-2002.
XX
XX  16-JAN-2002; 2002WO-US01395.
XX
XX  18-JAN-2001; 2001US-0765739.
XX
XX  (IDEX-) IDEX LAB INC.
XX
XX  Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX  WPI; 2002-599730/64.
XX
XX  New composition of matter comprising a polypeptide, useful in detecting
XX  the presence of antibodies to Ehrlichia canis or chaffeensis, or in
XX  detecting or quantifying the presence of Ehrlichia infection in mammals
XX
XX  Claim 1; Page 5; 29pp; English.
XX
XX  The invention relates to a composition of matter comprising a polypeptide
XX  isolated from Ehrlichia species. The composition can be used for
XX  detecting the presence of antibodies to Ehrlichia, comprising contacting
XX  one or more polypeptides with a test sample suspected of comprising
XX  antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
XX  complexes to form and detecting the complexes, where the detection of
XX  polypeptide/antibody complexes is an indication that antibodies to
XX  Ehrlichia are present in the test sample. The composition is useful for
XX  detecting or quantifying the presence of E. canis or E. chaffeensis
XX  infection in mammals. The polypeptides can be used to develop monoclonal
XX  and/or polyclonal antibodies that can be employed in assay systems and in
XX  the generation of chimeric antibodies for therapeutic use or other
XX  similar applications. This sequence represents an E. chaffeensis peptide
XX  fragment used in the composition of the invention.
XX
SQ  Sequence 20 AA;

Query Match          85.4%; Score 88; DB 23; Length 20;
Best Local Similarity 84.2%; Pred. No. 1.9e-07;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY  1 NNTTGVFGLKQDWDGSTIS 19
DB  |||||
1 NNTTGVFGLKQDWDGSAIS 19

RESULT 13
AAY06942
XX  AAY06942 standard; Protein; 256 AA.
XX
XX  AAY06942;
XX

```

DT 05-JUL-1999 (first entry)
XX E. chaffeensis p28 protein.
DE
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.
XX
OS Ehrlichia chaffeensis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX (OHIS) UNIV OHIO STATE.
PA
PI Ohashi N, Rikihisa Y;
XX
DR WPI: 1999-254290/21.
DR N-PSDB; AAX34742.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
PS Claim 18; Fig 1; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 256 AA;
Query Match 85.4%; Score 88; DB 20; Length 256;
Best Local Similarity 84.2%; Pred. No. 3.5e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQDWGSGTIS 19
Db 34 NTTGVFGLKQNWGSAIS 52
RESULT 14
ABG77966
ID ABG77966 standard; Protein; 256 AA.
XX
AC ABG77966;
XX
DT 15-NOV-2002 (first entry)
XX
DE Protein encoded by Ehrlichia chaffeensis p28 gene.
XX
KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection; p28; OMP-1.
XX
OS Ehrlichia chaffeensis.
XX
PN US2002120115-A1.
XX
PD 29-AUG-2002.
XX
PF 28-JAN-2002; 2002US-0059964.
XX
PR 19-MAY-1999; 99US-0314701.
XX
PA (RIKI/) RIKIHISA Y.
PA (OHAS/) OHASHI N.
XX
PI Rikihisa Y, Ohashi N;

XX WPI: 2002-618954/66.
DR N-PSDB; ABS63307.
XX
PT Isolated polynucleotide encoding an outer membrane protein of E. canis
PT or E. chaffeensis used in the diagnosis of infection -
XX
PS Disclosure; Fig 1; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents the Ehrlichia chaffeensis OMP-1 protein encoded
CC by the p28 gene.
XX
SQ Sequence 256 AA;
Query Match 85.4%; Score 88; DB 23; Length 256;
Best Local Similarity 84.2%; Pred. No. 3.5e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQDWGSGTIS 19
Db 34 NTTGVFGLKQNWGSAIS 52
RESULT 15
AAW51095
ID AAW51095 standard; Protein; 276 AA.
XX
AC AAW51095;
XX
DT 14-SEP-1998 (first entry)
XX
DE Ehrlichia chaffeensis VSA5 protein (partial sequence).
XX
KW MAP1 homologue; variable surface antigen; VSA5; rickettsia;
KW DNA vaccine.
XX
OS Ehrlichia chaffeensis.
XX
FH Key Location/Qualifiers
FT Peptide 1..25 /note= "putative signal peptide"
XX
PN WO9816554-A1.
XX
PD 23-APR-1998.
XX
PF 17-OCT-1997; 97WO-US19044.
XX
PR 17-OCT-1996; 96US-0733230.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
PI Nyika A, Rurangirwa FR;
XX
DR WPI: 1998-251232/22.
DR N-PSDB; AAV07179.
XX
PT Composition containing nucleic acid encoding rickettsial antigen -
PT useful for, e.g. stimulating protective immune response in humans or
PT animals
XX
PS Claim 3; Fig 2B; 39pp; English.
XX
CC This is the near full-length variable surface antigen VSA5 protein

of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid residues. The VSA5 amino acid sequence was deduced from a partial open reading frame (ORF5) of a genomic locus (see AAV071179) of E. chaffeensis that was obtained on the basis of homology to the major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This genomic locus included 5 ORFs encoding similar, but non-identical proteins (see AAW51091-95). A claimed composition comprises a nucleic acid (see AAV071176-82) encoding a polypeptide (see AAW51088-99) that elicits a protective immune response against a rickettsial pathogen. The nucleic acid is used, in human or veterinary medicine, in vaccines to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic polypeptides can be used diagnostically to detect antibodies associated with Ehrlichia infection (claimed).

SQ Sequence 276 AA;

Query Match 85.4%; Score 88; DB 19; Length 276;
Best Local Similarity 84.2%; Pred. No. 3.8e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDGSTIS 19
||| |||||:|||| ||
Db 59 NTTGVFGLKQNDGSAIS 77

RESULT 16

AAB36189
ID AAB36189 standard; Protein; 276 AA.

AC AAB36189;

DT 02-MAR-2001 (first entry)

DE Ehrlichia chaffeensis partial VSA5.

KW Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
major antigenic protein 1; antirickettsial; vaccine; gene therapy;
Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
3gdoorf3.

OS Ehrlichia chaffeensis.

PN WO200065063-A2.

PD 02-NOV-2000.

PF 21-APR-2000; 2000WO-US10886.

PR 22-APR-1999; 99US-0130725.

PA (UYFL) UNIV FLORIDA.

PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, McGuire TC;
Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allemen AR;

DR WPI: 2000-679675/66.

DR N-PSDB; AAC68706.

PT New polynucleotides useful as DNA vaccines for conferring immunity to
rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

PS Claim 3; Page 47; 63pp; English.

The present sequence shows a high degree of similarity to the major
antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
used in a vaccines to protect animals or humans against rickettsial
diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
protective against the rickettsial pathogen. The nucleic acid vaccines
can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.

CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
CC and 3gdoorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.

SQ Sequence 276 AA;

Query Match 85.4%; Score 88; DB 21; Length 276;

Best Local Similarity 84.2%; Pred. No. 3.8e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDGSTIS 19
||| |||||:|||| ||
Db 59 NTTGVFGLKQNDGSAIS 77

RESULT 17

AAU04199

ID AAU04199 standard; Protein; 276 AA.

AC AAU04199;

DT 23-OCT-2001 (first entry)

DE Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.

KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
infection; heartwater; diagnostic; variable surface antigen; VSA.

OS Ehrlichia chaffeensis.

PN US6251872-B1.

PD 26-JUN-2001.

PF 17-OCT-1997; 97US-0953326.

PR 17-OCT-1996; 96US-0733230.

PA (UYFL) UNIV FLORIDA.

PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
Rurangirwa FR, Mahan SM, Bowie MV, Allemen AR;

DR WPI: 2001-424487/45.

DR N-PSDB; AAS07578.

PT New MAP2 genes and polypeptides useful as vaccines for conferring
immunity to human and animal rickettsial diseases, e.g. heartwater, or
as molecular markers in nucleic acid analysis procedures

PS Example 3; Fig 2A-2B; 30pp; English.

CC The sequence represents the amino acid sequence of variable surface
antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.

SQ Sequence 276 AA;

Query Match 85.4%; Score 88; DB 22; Length 276;

Best Local Similarity 84.2%; Pred. No. 3.8e-06;

Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSGTIS 19
 ||| |||||:|||||
 Db 59 NTTGVFLKQKQWDGSAIS 77

RESULT 18

AAW51089
 ID AAW51089 standard; Protein; 280 AA.

XX AAW51089;

DT 14-SEP-1998 (first entry)

DE Ehrlichia chaffeensis major antigenic protein 1 (MAP1).

XX MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.

XX Ehrlichia chaffeensis.

OS W09816554-A1.

PN 23-APR-1998.

PD 17-OCT-1997; 97WO-US19044.

PF 17-OCT-1996; 96US-0733230.

PR (UYFL) UNIV FLORIDA.

PS Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;

PI Nyika A, Rurangirwa FR;

XX WPI; 1998-251232/22.

DR N-PSDB; AAV07177.

XX Composition containing nucleic acid encoding rickettsial antigen -
 useful for, e.g. stimulating protective immune response in humans or
 animals

PS Claim 3; Page 18-19; 39pp; English.

XX This polypeptide comprises the major antigen protein 1 gene (MAP1)
 of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see
 AAV07177). A claimed composition comprises a nucleic acid (see
 AAV07176-82) encoding a polypeptide (see AAW51088-99) that elicits a
 protective immune response against a rickettsial pathogen. The
 nucleic acid is used, in human or veterinary medicine, in vaccines
 to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
 species. The nucleic acid does not replicate in the host but
 remains episomal and capable of expressing polypeptide for at least
 19 mth. The Ehrlichia antigenic polypeptides can be used
 diagnostically to detect antibodies associated with Ehrlichia
 infection (claimed).

XX Sequence 280 AA;

Query Match 85.4%; Score 88; DB 19; Length 280;
 Best Local Similarity 84.2%; Pred. No. 3.9e-06;

Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSGTIS 19
 ||| |||||:|||||

Db 60 NTTGVFLKQKQWDGSAIS 78

RESULT 19

AAB36183
 ID AAB36183 standard; Protein; 280 AA.

XX AAB36183;

DT 02-MAR-2001 (first entry)

XX Ehrlichia chaffeensis MAP1.

XX Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
 vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
 4hworf1; 18hworf1; 3gdorf3.

OS Ehrlichia chaffeensis.

XX W0200065063-A2.

PN 02-NOV-2000.

PD 21-APR-2000; 2000WO-US10886.

PF 22-APR-1999; 99US-0130725.

PR (UYFL) UNIV FLORIDA.

PS Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allenan AR;

XX WPI; 2000-679675/66.

DR N-PSDB; AAC68700.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
 rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

PS Claim 3; Page 35-36; 63pp; English.

XX The present sequence is given in a specification relating to nucleic
 acid vaccines containing genes to protect animals or humans against
 rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
 sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 protective against the rickettsial pathogen. The vaccine comprises the
 major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
 (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
 driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
 ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1 and
 3gdorf3 may be used in therapeutic and diagnostic applications. The
 polypeptides are useful for detecting antibodies associated with
 infection by a rickettsial pathogen whilst the polynucleotides may be
 used to detect the presence of rickettsial nucleic acids.

XX Sequence 280 AA;

Query Match 85.4%; Score 88; DB 21; Length 280;

Best Local Similarity 84.2%; Pred. No. 3.9e-06;

Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSGTIS 19
 ||| |||||:|||||

Db 60 NTTGVFLKQKQWDGSAIS 78

RESULT 20

AAU04193

ID AAU04193 standard; Protein; 280 AA.

XX AAU04193;

XX 23-OCT-2001 (first entry)

DE Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.

XX Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
 infection; heartwater; diagnostic.

XX Ehrlichia chaffeensis.

```

PN US6251872-B1.
XX
XX
XX 26-JUN-2001.
XX
XX PF 17-OCT-1997; 97US-0953326.
XX
XX PR 17-OCT-1996; 96US-0733230.
XX
XX (UYFL ) UNIV FLORIDA.
XX
XX Barbet AF, Ganta RR, McGuire TC, BurrIDGE MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
XX WPI: 2001-424487/45.
DR N-PSDB; AAS07576.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures -
XX
XX Disclosure; Column 15-17; 30pp; English.
XX
XX The sequence represents the amino acid sequence of major antigenic
CC protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
CC polypeptides are useful as vaccines for conferring immunity to rickettsia
CC infection, including Cowdria ruminantium causing heartwater. The MAP
CC polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.
XX
XX Sequence 280 AA;
XX
XX Query Match 85.4%; Score 88; DB 22; Length 280;
XX Best Local Similarity 84.2%; Pred. No. 3.9e-06;
XX Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 NTTGVFGLKQWDGSGTIS 19
XX ||| |||||:|||| ||
XX 60 NTTGVFGLKQWDGSAIS 78
XX
XX RESULT 21
XX AAY06943
XX ID AAY06943 standard; Protein; 281 AA.
XX
XX AC AAY06943;
XX
XX DT 05-JUL-1999 (first entry)
XX
XX DE E. chaffeensis OMP-1 protein.
XX
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX KW detection; dog.
XX
XX OS Ehrlichia chaffeensis.
XX
XX PN WO9913720-A1.
XX
XX PD 25-MAR-1999.
XX
XX PF 18-SEP-1998; 98WO-US19600.
XX
XX PR 19-SEP-1997; 97US-0059353.
XX
XX PA (OHLS ) UNIV OHIO STATE.
XX
XX PI Ohashi N, Rikihisa Y;
XX
XX WPI: 1999-254290/21.
DR N-PSDB; AAX34743.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
XX Disclosure; Fig 3B; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY05959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 281 AA;
XX
XX Query Match 85.4%; Score 88; DB 20; Length 281;
XX Best Local Similarity 84.2%; Pred. No. 3.9e-06;
XX Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 NTTGVFGLKQWDGSGTIS 19
XX ||| |||||:|||| ||
XX 59 NTTGVFGLKQWDGSAIS 77
XX
XX RESULT 22
XX ABG77935
XX ID ABG77935 standard; Protein; 281 AA.
XX
XX AC ABG77935;
XX
XX DT 15-NOV-2002 (first entry)
XX
XX DE Ehrlichia chaffeensis outer membrane protein (OMP) #1.
XX
XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
XX OS Ehrlichia chaffeensis.
XX
XX PN US2002120115-A1.
XX
XX PD 29-AUG-2002.
XX
XX PF 28-JAN-2002; 2002US-0059964.
XX
XX PR 19-MAY-1999; 99US-0314701.
XX
XX PA (RIKI/) RIKIHISA Y.
XX PA (OHAS/) OHASHI N.
XX
XX PI Rikihisa Y, Ohashi N;
XX
XX DR WPI: 2002-618954/66.
XX DR N-PSDB; ABS63276.
XX
XX Isolated polynucleotide encoding an outer membrane protein of E. canis
PT or E. chaffeensis used in the diagnosis of infection -
XX
XX Claim 14; Fig 3B; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX

```

```

SQ Sequence 281 AA;
Query Match 85.4%; Score 88; DB 23; Length 281;
Best Local Similarity 84.2%; Pred. No. 3.9e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
   III|IIIIII|IIII|II
Db 59 NTTGVFGLKQNWGSAIS 77

RESULT 23
AAU96105
ID AAU96105 standard; Protein; 281 AA.
XX AC
XX AAU96105;
XX 02-JUL-2002 (first entry)
XX DE
XX Ehrlichia chaffeensis P28.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX OS Ehrlichia chaffeensis.
XX PN WO200222782-A2.
XX PD 21-MAR-2002.
XX PF 12-SEP-2001; 2001WO-US28759.
XX PR 12-SEP-2000; 2000US-0660587.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X, McBride JW;
XX WPI; 2002-351882/38.
XX DE New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX Example 3; Figure 3; 106pp; English.
XX CC The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
SQ Sequence 281 AA;
Query Match 85.4%; Score 88; DB 23; Length 281;
Best Local Similarity 84.2%; Pred. No. 3.9e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
   III|IIIIII|IIII|II
Db 59 NTTGVFGLKQNWGSAIS 77

RESULT 24
AAU73418
ID AAU73418 standard; Protein; 281 AA.
XX AC
XX AAU73418;
XX 12-MAR-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein P28-19.
```

```

XX Ehrlichia; outer membrane protein; p28; antibiotic; vaccine.
XX Ehrlichia chaffeensis.
XX WO200183699-A2.
XX 08-NOV-2001.
XX 01-MAY-2001; 2001WO-US13997.
XX 01-MAY-2000; 2000US-201035P.
XX (RERE-) RES DEV FOUND.
XX Walker DH, Yu X;
XX WPI; 2002-066527/09.
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis -
XX Disclosure; Figure 2; 97pp; English.
XX CC The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
SQ Sequence 281 AA;
Query Match 85.4%; Score 88; DB 23; Length 281;
Best Local Similarity 84.2%; Pred. No. 3.9e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
   III|IIIIII|IIII|II
Db 59 NTTGVFGLKQNWGSAIS 77

RESULT 25
AAU06962
ID AAY06962 standard; Protein; 280 AA.
XX AC AAY06962;
XX 05-JUL-1999 (first entry)
XX DE E. canis P30-2 protein.
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX OS Ehrlichia canis.
XX PN WO9913720-A1.
XX 25-MAR-1999.
XX 18-SEP-1998; 98WO-US19600.
XX 19-SEP-1997; 97US-0059353.
XX (OHIS ) UNIV OHIO STATE.
XX Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
XX N-PSDB; AAX34762.
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
```

CC	Canine ehrlichiosis, also known as canine tropical pancytopenia, is a tick-borne rickettsial disease of dogs. ECA28-1 is conserved amongst different strains of E. canis and hence useful for serodiagnosis of canine ehrlichiosis. The present sequence is a E. canis ECA28SA3 30-kDa protein which is post-translationally modified to a mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX	
SQ	Sequence 280 AA;
	Query Match 77.7%; Score 80; DB 21; Length 280; Best Local Similarity 73.7%; Pred. No. 7.9e-05; Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY	1 NTTCGVFLGKQMDGSTIS 19 I:I I:IIIII II:I III
Db	59 NSTGVFGFKHDWNGGTIS 77
RESULT 27	
ID	ABG77953
ID	ABG77953 standard; Protein; 280 AA.
AC	XX
AC	ABG77953;
DT	15-NOV-2002 (first entry)
DE	Ehrlichia canis outer membrane protein (P30F) #4.
XX	Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
KW	Ehrlichia canis.
OS	US2002120115-Al.
XX	
PN	XX
XX	29-AUG-2002.
PD	
Pf	28-JAN-2002; 200ZUS-0059964.
XX	
PR	19-MAY-1999; 99US-0314701.
XX	
PA	{RIKIHI} RIKIHISA Y. (OHAS/) OHASHI N.
PI	Rikihisa Y, Ohashi N;
XX	
DR	WPI; 2002-618954/66. N-PSDB; ABS63294.
XX	
PT	Isolated polynucleotide encoding an outer membrane protein of E.canis or E.chaffeensis used in the diagnosis of infection -
PS	Claim 10; Fig 22B; 49pp; English.
XX	
CC	The invention relates to an isolated polynucleotide encoding an outer membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used in the diagnosis of infection. An infection such as human ehrlichiosis or canine ehrlichiosis can be diagnosed by providing a serum sample from the patient, providing a polypeptide or mixture of polypeptides, contacting the sample with the polypeptide and assaying for the formation of a complex between antibodies in the serum sample and the polypeptide, where formation of a complex is indicative of infection with E. chaffeensis. This sequence represents an Ehrlichia outer membrane protein of the invention.
XX	
SQ	Sequence 280 AA;
	Query Match 77.7%; Score 80; DB 23; Length 280; Best Local Similarity 73.7%; Pred. No. 7.9e-05; Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY	1 NTTCGVFLGKQMDGSTIS 19 I:I I:IIIII II:I III
Db	59 NSTGVFGFKHDWNGGTIS 77

```

XX PF 16-JAN-2002; 2002WO-US01395.
XX PR 18-JAN-2001; 2001US-0765739.
XX PA (IDEX-) IDEXX LAB INC.
XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX WPI; 2002-599730/64.
XX DR New composition of matter comprising a polypeptide, useful in detecting
XX PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
XX PT detecting or quantifying the presence of Ehrlichia infection in mammals
XX PS Claim 1; Page 5; 29pp; English.
XX CC The invention relates to a composition of matter comprising a polypeptide
XX CC isolated from Ehrlichia species. The composition can be used for
XX CC detecting the presence of antibodies to Ehrlichia, comprising contacting
XX CC one or more polypeptides with a test sample suspected of comprising
XX CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
XX CC complexes to form and detecting the complexes, where the detection of
XX CC polypeptide/antibody complexes is an indication that antibodies to
XX CC Ehrlichia are present in the test sample. The composition is useful for
XX CC detecting or quantifying the presence of E. canis or E. chaffeensis
XX CC infection in mammals. The polypeptides can be used to develop monoclonal
XX CC and/or polyclonal antibodies that can be employed in assay systems and in
XX CC the generation of chimeric antibodies for therapeutic use or other
XX CC similar applications. This sequence represents an E. chaffeensis peptide
XX CC fragment used in the composition of the invention.
XX SQ Sequence 19 AA;

Query Match 71.8%; Score 74; DB 23; Length 19;
Best Local Similarity 68.4%; Pred. No. 3.4e-05;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGGSTIS 19
   ||| ||||:|||| ||
DB 1 NTTGVFGIEQDWDRCVIS 19

RESULT 30
AAW51092
ID AAW51092 standard; Protein; 286 AA.
XX AC
XX AC AAW51092;
XX DT 14-SEP-1998 (first entry)
XX DE Ehrlichia chaffeensis VSA2 protein.
XX DE MAP1 homologue; variable surface antigen; VSA2; rickettsia;
XX KW DNA vaccine.
XX OS Ehrlichia chaffeensis.
XX FH Key Location/Qualifiers
XX FT Peptide 1..25
XX FT /note= "putative signal peptide"
XX PN WO9816554-A1.
XX XX
XX PD 23-APR-1998.
XX PF 17-OCT-1997; 97WO-US19044.
XX XX
XX PR 17-OCT-1996; 96US-0733230.
XX XX
XX PA (UYFL ) UNIV FLORIDA.
XX XX

RESULT 28
AAU96102
ID AAU96102 standard; Protein; 280 AA.
XX AC
XX AC AAU96102;
XX DT 02-JUL-2002 (first entry)
XX DE Ehrlichia canis p28-6.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX OS Ehrlichia canis.
XX PN WO200222782-A2.
XX PD
XX PD 21-MAR-2002.
XX PF 12-SEP-2001; 2001WO-US28759.
XX PR 12-SEP-2000; 2000US-0660587.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X, McBride JW;
XX WPI; 2002-351882/38.
XX N-PSDB; ABK68854.
XX CC New recombinant homologous 28 kilodalton immunodominant protein from
XX CC Ehrlichia canis, useful for treating Ehrlichia canis infections
XX PS Claim 16; Figure 7; 106pp; English.
XX CC The invention relates to a recombinant homologous 28 kDa immunodominant
XX CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
XX CC dispersed in a pharmaceutically acceptable carrier, is useful for
XX CC inhibiting E. canis infection in a subject. (I) is useful in the
XX CC development of vaccines and serodiagnostics that are particularly
XX CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
XX CC represent the 28-kDa antigen amino acid sequences of the invention.
XX SQ Sequence 280 AA;

Query Match 77.7%; Score 80; DB 23; Length 280;
Best Local Similarity 73.7%; Pred. No. 7.9e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGGSTIS 19
   ||:||||||| ||:||||
DB 59 NSTGVFGLKHDWNGGTIS 77

RESULT 29
ABG30747
ID ABG30747 standard; Peptide; 19 AA.
XX AC
XX AC ABG30747;
XX DT 21-OCT-2002 (first entry)
XX DE Ehrlichia chaffeensis peptide fragment #3.
XX KW Antibody detection; monoclonal antibody; polyclonal antibody.
XX OS Ehrlichia chaffeensis.
XX PN WO200257794-A2.
XX PD 25-JUL-2002.

```

PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 XX Nyika A, Rurangirwa FR;
 DR WPI: 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX
 PT Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 PS Claim 3; Fig 2A; 39pp; English.
 XX
 CC This is the full-length variable surface antigen VSA2 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF2) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 XX Sequence 286 AA;
 SQ
 Query Match 71.8%; Score 74; DB 19; Length 286;
 Best Local Similarity 68.4%; Pred. No. 0.00077;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NTTGVEGLKQDWGSGTIS 19
 Db |||||:|||||
 60 NTTGVFGIEQDWRCVIS 78
 RESULT 31
 AAY06946
 ID AAY06946 standard; Protein; 286 AA.
 AC AAY06946;
 XX
 DT 05-JUL-1999 (first entry)
 DE E. chaffeensis OMP-ID protein.
 XX
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO9913720-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 18-SEP-1998; 98WO-US19600.
 XX
 PR 19-SEP-1997; 97US-0059353.
 XX
 PA (OHIS) UNIV OHIO STATE.
 XX
 PI Ohashi N, Rikihisa Y;
 XX
 DR WPI: 1999-254290/21.
 DR N-PSDB; AAX34746.
 XX
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 XX
 PS Claim 14; Fig 6B; 55pp; English.
 XX

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 XX Sequence 286 AA;
 SQ

Query Match 71.8%; Score 74; DB 20; Length 286;
 Best Local Similarity 68.4%; Pred. No. 0.00077;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVEGLKQDWGSGTIS 19
 Db |||||:|||||
 60 NTTGVFGIEQDWRCVIS 78

RESULT 32
 AAB36186

ID AAB36186 standard; Protein; 286 AA.

AC AAB36186;

DT 02-MAR-2001 (first entry)

DE Ehrlichia chaffeensis partial VSA2.

KW Ehrlichia chaffeensis; VSA2; variable surface antigen 2; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.

OS Ehrlichia chaffeensis.

PN WO200065063-A2.

PD 02-NOV-2000.

PF 21-APR-2000; 2000WO-US10886.

PR 22-APR-1999; 99US-0130725.

PA (UYFL) UNIV FLORIDA.

PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbl BH, Whitmire WW, Allenman AR;
 XX
 DR WPI: 2000-679675/66.
 DR N-PSDB; AAC68703.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
 XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 XX

Claim 3; Page 43-44; 63pp; English.

XX The present sequence shows a high degree of similarity to the major
 XX antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 XX used in a vaccine to protect animals or humans against rickettsial
 XX diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 XX Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 XX protective against the rickettsial pathogen. The nucleic acid vaccines
 XX can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 XX Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 XX and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 XX polypeptides are useful for detecting antibodies associated with
 XX infection by a rickettsial pathogen whilst the polynucleotides may be
 XX used to detect the presence of rickettsial nucleic acids.

XX Sequence 286 AA;
 XX

Query Match 71.8%; Score 74; DB 21; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00077;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSGTIS 19
||| ||||:|||| ||
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 33

AAU04196
ID AAU04196 standard; Protein: 286 AA.

XX AC AAU04196;

XX DT 23-OCT-2001 (first entry)

XX DE Variable surface antigen 2 (VSA2) from Ehrlichia chaffeensis.

XX KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
XX KW infection; heartwater; diagnostic; variable surface antigen; VSA.

XX OS Ehrlichia chaffeensis.

XX PN USG251872-B1.

XX PD 26-JUN-2001.

XX PF 17-OCT-1997; 97US-0953326.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UVFL) UNIV FLORIDA.

PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;

PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

DR WPI; 2001-424487/45.

DR N-PSDB; AAS07578.

XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures -

XX Example 3; Fig 2A-2B; 30pp; English.

PS The sequence represents the amino acid sequence of variable surface
CC antigen 2 (VSA2) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.

XX SQ Sequence 286 AA;

Query Match 71.8%; Score 74; DB 22; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00077;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSGTIS 19
||| ||||:|||| ||
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 34

ABG77938
ID ABG77938 standard; Protein: 286 AA.

XX AC ABG77938;

XX DT 15-NOV-2002 (first entry)

XX DE Ehrlichia chaffeensis outer membrane protein (OMP) #4.

XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

XX OS Ehrlichia chaffeensis.

XX PN US2002120115-A1.

XX PD 29-AUG-2002.

XX PF 28-JAN-2002; 2002US-0059964.

XX PR 19-MAY-1999; 99US-0314701.

XX PA (RIKI/) RIKIHISA Y.

XX PA (OHAS/) OHASHI N.

XX PI Rikihisa Y, Ohashi N;

XX DR WPI; 2002-618954/66.

XX DR N-PSDB; ABS63279.

XX PT Isolated polynucleotide encoding an outer membrane protein of E.canis
or E.chaffeensis used in the diagnosis of infection -

XX PS Disclosure; Fig 6B; 49pp; English.

XX CC The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.

XX SQ Sequence 286 AA;

Query Match 71.8%; Score 74; DB 23; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00077;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSGTIS 19
||| ||||:|||| ||
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 35

AAU96108
ID AAU96108 standard; Protein: 286 AA.

XX AC AAU96108;

XX DT 02-JUL-2002 (first entry)

XX DE Ehrlichia chaffeensis OMP-ID.

XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
antibacterial.

XX OS Ehrlichia chaffeensis.

XX PN WO200222782-A2.

PD 21-MAR-2002.
 XX
 XX 12-SEP-2001; 2001WO-US28759.
 XX
 PR 12-SEP-2000; 2000US-0660587.
 XX
 PA (RERE-) RES DEV FOUND.
 XX
 PI Walker DH, Yu X, McBride JW;
 XX
 DR WPI; 2002-351882/38.
 XX
 PT New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
 XX
 PS Example 3; Figure 3; 106pp; English.
 XX
 CC The invention relates to a recombinant homologous 28 kDa immunodominant
 CC protein, P28 (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX
 SQ Sequence 286 AA;

Query Match 71.8%; Score 74; DB 23; Length 286;
 Best Local Similarity 68.4%; Pred NO; 0.00077;
 Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 NTTTGVEGLKQDWDGDTIS 19
 ||| ||||:||||
 Db 60 NTTVGVEGIEQDWDRCVIS 78

Search completed: October 6, 2003, 07:48:02
 Job time : 34.6045 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 27.1642 Seconds
(without alignments)
189,995 Million cell updates/sec

Title: US-09-765-739A-2

Perfect score: 110

Sequence: 1 NTTGVFGLKQDWDGATIKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052504 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SPTREMBL_23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	288	2	Q9ZGJ2
2	96	87.3	280	2	Q52107
3	86	78.2	246	2	Q9RH35
4	86	78.2	275	2	Q93DD4
5	86	78.2	276	2	Q93DD1
6	86	78.2	276	2	Q85817
7	86	78.2	276	2	Q85GU0
8	82	74.5	281	2	Q93DD2
9	82	74.5	281	2	Q9AC19
10	81	73.6	280	2	Q93DD3
11	81	73.6	280	2	Q9ZGM9
12	81	73.6	280	2	Q85816
13	77	70.0	280	2	Q9ADV3
14	77	70.0	280	2	Q9F473
15	76	69.1	276	2	Q9F475
16	71.5	65.0	287	2	Q8GGU3

Q8GGU2	ehrlchia c
Q8G921	ehrlchia c
Q52105	ehrlchia c
Q9F472	ehrlchia c
Q9RA8	ehrlchia c
Q9RA7	ehrlchia c
Q9R3J3	ehrlchia c
Q9RA6	ehrlchia c
Q9RA9	ehrlchia c
Q9RA5	ehrlchia c
Q9ZGJ1	ehrlchia c
Q8G9P3	ehrlchia c
Q8G948	ehrlchia c
Q8GGU1	ehrlchia c
Q8G8W7	ehrlchia c
Q9AF98	ehrlchia c
Q9AF98	ehrlchia c
Q93E57	ehrlchia c
Q93E52	ehrlchia c
Q93E57	ehrlchia c
Q93E56	ehrlchia c
Q46324	ehrlchia c
Q93E64	ehrlchia c
Q46332	ehrlchia c
Q9AEU3	ehrlchia c
Q9AF99	ehrlchia c
Q93E59	ehrlchia c
Q93E60	ehrlchia c
Q93E53	ehrlchia c
Q46329	ehrlchia c
Q9R425	ehrlchia c
Q46331	ehrlchia c
Q46333	ehrlchia c
Q46330	ehrlchia c
Q9AMF6	ehrlchia c
Q93E54	ehrlchia c
Q52106	ehrlchia c
Q8G8Q5	ehrlchia c
Q52104	ehrlchia c
Q8GGU5	ehrlchia c
Q8G8J3	ehrlchia c
Q9AF41	ehrlchia c
Q46327	ehrlchia c
Q93E55	ehrlchia c
Q93E58	ehrlchia c
Q93E62	ehrlchia c
Q9S6H1	ehrlchia c
Q46328	ehrlchia c
Q9S6H0	ehrlchia c
Q9AF40	ehrlchia c
Q91964	ehrlchia c
Q8X4H5	ehrlchia c
Q8X2T1	ehrlchia c
Q93E61	ehrlchia c
Q93E63	ehrlchia c
Q85360	ehrlchia c
Q9E2L0	ehrlchia c
Q9E2K9	ehrlchia c
Q9E2L1	ehrlchia c
Q9E2K6	ehrlchia c
Q9F474	ehrlchia c
Q8HA41	ehrlchia c
Q97H03	ehrlchia c
Q06352	ehrlchia c
Q98SL3	ehrlchia c
Q9H201	ehrlchia c
Q9K7U6	ehrlchia c
Q76154	ehrlchia c
Q9E2K8	ehrlchia c
Q9E2K5	ehrlchia c
Q9FKI3	ehrlchia c
Q818M9	ehrlchia c
Q8N13	ehrlchia c
Q95P44	ehrlchia c
Q9GPG2	ehrlchia c

90 Q8u9m0 agrobacteri
91 Q8imb4 drosophila
92 Q9w3r8 drosophila
93 Q8p3y5 xanthomonas
94 Q8x5n0 escherichia
95 Q8fci5 escherichia
96 Q9kpt3 vibrio chol
97 Q8na86 homo sapien
98 Q8fnr9 corynebacte
99 Q9w5y2 drosophila
100 Q9u458 drosophila

ALIGNMENTS

RESULT 1

Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
ID Q9ZGJ2
AC Q9ZGJ2
DT 01-MAY-1999 (TREMREL. 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE 30-kDa major outer membrane protein (P28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
CX NCBI_TaxID=944;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis";
RL Gene 254:245-252(2000).
DR EMBL; AF078553; AAC68667.1; -;
DR EMBL; AF082744; AGL4362.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BF5E CRC64;

Query Match 100.0%; Score 110; DB 2; Length 288;
Best Local Similarity 100.0%; Pred. No. 3.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTTGVFLKQDWGATIKD 20
|||||
Db 60 NTTTGVFLKQDWGATIKD 79

RESULT 2

Q52107 PRELIMINARY; PRT; 280 AA.
ID Q52107
AC Q52107
DT 01-JUN-1998 (TREMREL. 06, Created)
DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Major outer membrane protein OMP-1F (28kDa outer membrane protein gene
DE 18).
GN OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
CX NCBI_TaxID=945;
FN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulisoma C.R., Barbet A.F., Mahan S.M., Burrridge M.J.,
RA Allenman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02940.1; -;
DR EMBL; AF479833; AAO12931.1; -;
DR EMBL; AF479834; AAO12937.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;

Query Match 87.3%; Score 96; DB 2; Length 280;

Best Local Similarity 94.4%; Pred. No. 5.9e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTTGVFLKQDWGATIKD 18
|||||
Db 60 NTTTGVFLKQDWGATIKD 77

RESULT 3

Q9RH35 PRELIMINARY; PRT; 246 AA.
ID Q9RH35
AC Q9RH35
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
DE Outer membrane protein p28 (Fragment).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

```
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St.Vincent;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 78.2%; Score 86; DB 2; Length 246;
Best Local Similarity 83.3%; Pred. No. 1.9e-05;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 18
Db 29 NTTAGVFLKQDWDGSAI 46

RESULT 4
Q93DD4 PRELIMINARY; PRT; 275 AA.
AC Q93DD4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V2;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393389; AAL12919.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 29974 MW; 2ECCF2F98B2E9D9 CRC64;

Query Match 78.2%; Score 86; DB 2; Length 275;
Best Local Similarity 83.3%; Pred. No. 2.2e-05;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 18
Db 59 NTTAGVFLKQDWDGSAI 76

RESULT 5
Q93DD1 PRELIMINARY; PRT; 276 AA.
AC Q93DD1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393389; AAL12919.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30028 MW; 2D7143AFCBF2EBE CRC64;

Query Match 78.2%; Score 86; DB 2; Length 276;
Best Local Similarity 83.3%; Pred. No. 2.2e-05;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGATI 18
Db 59 NTTAGVFLKQDWDGSAI 76

RESULT 6
Q85817 PRELIMINARY; PRT; 276 AA.
AC Q85817;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28 kDa outer membrane protein).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V9;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Chattanooga, West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF077734; AAC31547.1; -
DR EMBL: AF393395; AAL12925.1; -
DR EMBL: AY117397; AAM77032.1; -
DR EMBL: AF479835; AAO12943.1; -
DR EMBL: AF479835; AAO12948.1; -
DR EMBL: AF479837; AAO12953.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 78.2%; Score 86; DB 2; Length 276;
Best Local Similarity 83.3%; Pred. No. 2.2e-05;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 NNTTGVFGLKQDWDGATI 18
AC III | | | | | | | | | | | |
Db 59 NNTAGVFLKQDWDGSAI 76

RESULT 7

Q8GGUO ID Q8GGUO PRELIMINARY; PRT; 276 AA.
AC Q8GGUO;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 19.
OS Ehrlichia chaffeensis
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wakulla;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479838; AAO12958.1;
SQ SEQUENCE 276 AA; 30045 MW; 8E3C9719D3C67A64 CRC64;

Query Match 78.2%; Score 86; DB 2; Length 276;
Best Local Similarity 83.3%; Pred. No. 2.2e-05;
Matches 15; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NNTTGVFGLKQDWDGATI 18
AC III | | | | | | | | | | | |
Db 59 NNTAGVFLKQDWDGSAI 76

RESULT 8

Q93DD2 ID Q93DD2 PRELIMINARY; PRT; 281 AA.
AC Q93DD2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V6;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393392; AAL12922.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30388 MW; B54DBB55AC839A9A CRC64;

Query Match 74.5%; Score 82; DB 2; Length 281;
Best Local Similarity 70.0%; Pred. No. 9.4e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NNTTGVFGLKQDWDGATIKD 20
AC III | | | | | | | | | | | |
Db 59 NNTAGVFLKQDWDGSAISN 78

RESULT 9

Q9AC19

Q9AC19 ID Q9AC19 PRELIMINARY; PRT; 281 AA.
AC Q9AC19;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Major outer membrane protein P28 (28 kDa outer membrane protein).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RA MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RA MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer Membrane Protein Multigene Family in Ehrlichia canis and E. chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V1;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Lithonia, Arkansas, and Osciola;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J., Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of the tribe Ehrlichiae";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; U72291; AKK28673.1;
DR EMBL; AF393388; AAL12918.1;
DR EMBL; AY117396; AAM77031.1;
DR EMBL; AF479833; AAO12932.1;
DR EMBL; AF479833; AAO12932.1;
DR EMBL; AF479833; AAO12938.1;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 74.5%; Score 82; DB 2; Length 281;
Best Local Similarity 70.0%; Pred. No. 9.4e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NNTTGVFGLKQDWDGATIKD 20
AC III | | | | | | | | | | | |
Db 59 NNTAGVFLKQDWDGSAISN 78

RESULT 10

Q93DD3

```
.ID Q93DD3 PRELIMINARY; PRT; 280 AA.
AC Q93DD3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V5;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393391; AAL12921.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30372 MW; C7B8C8710BC167E9 CRC64;

Query Match 73.6%; Score 81; DB 2; Length 280;
Best Local Similarity 77.8%; Pred. No. 0.00013;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGATI 18
Db :|| ||||| ||||| ||
59 STTAGVFLKQDWGSAI 76

RESULT 11
Q9ZGM9 PRELIMINARY; PRT; 280 AA.
AC Q9ZGM9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RA Yu X.-J., Walker D.H.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077732; AAC31545.1; -.
DR EMBL; AF393394; AAL12924.1; -.
DR EMBL; AF393390; AAL12920.1; -.
DR EMBL; AF479839; AAO12964.1; -.
DR EMBL; AF479840; AAO12970.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 73.6%; Score 81; DB 2; Length 280;
Best Local Similarity 77.8%; Pred. No. 0.00013;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGATI 18
Db :|| ||||| ||||| ||
59 STTAGVFLKQDWGSAI 76

RESULT 13
Q9ADV3 PRELIMINARY; PRT; 280 AA.
AC Q9ADV3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein p30-2.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
```

```
AC O85816;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28kDa outer membrane protein gene
DE 19).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V8, and V4;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077732; AAC31545.1; -.
DR EMBL; AF393394; AAL12924.1; -.
DR EMBL; AF393390; AAL12920.1; -.
DR EMBL; AF479839; AAO12964.1; -.
DR EMBL; AF479840; AAO12970.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 73.6%; Score 81; DB 2; Length 280;
Best Local Similarity 77.8%; Pred. No. 0.00013;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGATI 18
Db :|| ||||| ||||| ||
59 STTAGVFLKQDWGSAI 76

RESULT 13
Q9ADV3 PRELIMINARY; PRT; 280 AA.
AC Q9ADV3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major outer membrane protein p30-2.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
```

RESULT 15		PRELIMINARY		PRT	276 AA
ID	Q9F475				
AC	Q9F475				
DC	01-MAR-2001 (TReMBLrel. 16, Created)				
DT	01-JUN-2001 (TReMBLrel. 16, Last sequence update)				
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)				
DE	P28-3 (Major outer membrane protein P30-4).				
GN	P28-3 OR P30-4.				
OS	Ehrlichia canis.				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;				
OC	Anaplasmataceae; Ehrlichia.				
OX	NCBI_TaxID=944;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Jake;				
RC	MEDLINE=99242757; PubMed=10225842;				
RA	McBride J.W., Yu, X.J., Walker D.H.;				
RT	"Molecular cloning of the gene for a conserved major immunoreactive				
RT	28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic				
RT	antigen.";				
RL	Clin. Diagn. Lab. Immunol. 6:392-399(1999).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Jake;				
RC	MEDLINE=20432107; PubMed=10974556;				
RA	McBride J.W., Yu X.J., Walker D.H.;				
RT	"A conserved, transcriptionally active p28 multigene locus of				
RT	Ehrlichia canis.";				
RL	Gene 254:245-252(2000).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Oklahoma;				
RC	MEDLINE=98371112; PubMed=9705412;				
RA	Ohashi N., Unver A., Zhi N., Rikihisa Y.;				
RT	"Cloning and characterization of multigenes encoding the				
RT	immunodominant 30-kilodalton major outer membrane proteins of				
RT	Ehrlichia canis, and application of the recombinant protein for				
RT	serodiagnosis.";				
RL	J. Clin. Microbiol. 36:2671-2680(1998).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Oklahoma;				
RC	MEDLINE=21153566; PubMed=11254561;				
RA	Ohashi N., Rikihisa Y., Unver A.;				
RT	"Analysis of Transcriptionally Active Gene Clusters of Major Outer				
RT	Membrane Protein Multigene Family in Ehrlichia canis and E.				
RT	chafeensis.";				
RL	Infect. Immun. 69:2083-2091(2001).				
DR	EMBL; AF082744; AAG14358.1; -				
DR	EMBL; AF324792; AAK3131.1; -				
DR	EMBL; AF078553; AAK28697.1; -				
DR	InterPro; IPR002566; Surface_Ag_msp4.				
DR	Pfam; PF01617; Surface_Ag.2.1.				
SQ	SEQUENCE 276 AA; 30659 MW; CE51AB37D17AF3A4 CRC64;				
Query Match 69,13; Score 76; DB 2; Length 276;					
Best Local Similarity 65,08; Pred. No. 0.00081;					
Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps					
QY	1 NTTTGVFLKODWGATIKD 20				
DB	60 NITTFGLKESWTGGIILD 79				
RESULT 16					
ID	Q8GGU3				
AC	Q8GGU3				
DC	01-MAR-2003 (TReMBLrel. 23, Created)				
DT	01-MAR-2003 (TReMBLrel. 23, Last sequence update)				
DT	01-MAR-2003 (TReMBLrel. 23, Last annotation update)				

```
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St. Vincent.
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479837; AAO12951.1; -.
SQ SEQUENCE 287 AA; 31504 MW; A9592249C83695B2 CRC64;

Query Match 65.0%; Score 71.5; DB 2; Length 287;
Best Local Similarity 56.0%; Pred. No. 0.0043;
Matches 14; Conservative 3; Mismatches 3; Indels 5; Gaps 1;

QY 1 NTTGVFGLKQDWD-----GATIKD 20
DB 60 NTTGVFGIEQDWDRCVISRATLSD 84

RESULT 17
O8GGU2
ID O8GGU2 PRELIMINARY; PRT; 291 AA.
AC O8GGU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479838; AAO12956.1; -.
SQ SEQUENCE 291 AA; 32005 MW; 22D35EDED283195A CRC64;

Query Match 65.0%; Score 71.5; DB 2; Length 291;
Best Local Similarity 56.0%; Pred. No. 0.0044;
Matches 14; Conservative 3; Mismatches 3; Indels 5; Gaps 1;

QY 1 NTTGVFGLKQDWD-----GATIKD 20
DB 60 NTTGVFGIEQDWDRCVISRATLSD 84

RESULT 18
Q8G921
ID Q8G921 PRELIMINARY; PRT; 291 AA.
AC Q8G921;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=West Paces, and Heartland;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479835; AAO12941.1; -.
DR EMBL: AF479836; AAO12946.1; -.
SQ SEQUENCE 291 AA; 31991 MW; 22D35A2202831369 CRC64;

Query Match 65.0%; Score 71.5; DB 2; Length 291;
Best Local Similarity 56.0%; Pred. No. 0.0044;
Matches 14; Conservative 3; Mismatches 3; Indels 5; Gaps 1;

QY 1 NTTGVFGLKQDWD-----GATIKD 20
DB 60 NTTGVFGIEQDWDRCVISRATLSD 84

RESULT 19
O52105
ID O52105 PRELIMINARY; PRT; 286 AA.
AC O52105;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OMP-1D (28kDa outer membrane protein gene 16).
OS OMP-1D.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allaman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: U72291; AAC02938.1; -.
DR EMBL: AF479833; AAC26718.1; -.
DR EMBL: AF479834; AAO12935.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag2; 1.
SQ SEQUENCE 286 AA; 31509 MW; F145A79270F386BE CRC64;

Query Match 64.5%; Score 71; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 0.0051;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGATI 18
DB 60 NTTGVFGIEQDWDRCVI 77
```

```

Query Match      63.6%; Score 70; DB 2; Length 278;
Best Local Similarity 70.6%; Pred. No. 0.0072;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 TTTGVFLKQDWDGATI 18
DB      61 STVGVEGLKHDWDGSP 77

RESULT 20
Q9F472 ID Q9F472 PRELIMINARY; PRT; 278 AA.
AC Q9F472;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-7.
GN P28-7.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF082744; AAC64550.2; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30455 MW; 2411CAAB4C56CA74 CRC64;

Query Match      63.6%; Score 70; DB 2; Length 278;
Best Local Similarity 70.6%; Pred. No. 0.0072;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 TTTGVFLKQDWDGATI 18
DB      61 STVGVEGLKHDWDGSP 77

RESULT 21
Q9R8A8 ID Q9R8A8 PRELIMINARY; PRT; 278 AA.
AC Q9R8A8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082746; AAC64552.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match      63.6%; Score 70; DB 2; Length 278;
Best Local Similarity 70.6%; Pred. No. 0.0072;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 TTTGVFLKQDWDGATI 18
DB      61 STVGVEGLKHDWDGSP 77

RESULT 22
Q9R8A7 ID Q9R8A7 PRELIMINARY; PRT; 278 AA.
AC Q9R8A7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Demon;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082747; AAC64553.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match      63.6%; Score 70; DB 2; Length 278;
Best Local Similarity 70.6%; Pred. No. 0.0072;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 TTTGVFLKQDWDGATI 18
DB      61 STVGVEGLKHDWDGSP 77

RESULT 23
Q9R3J3 ID Q9R3J3 PRELIMINARY; PRT; 278 AA.
AC Q9R3J3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Florida;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082750; AAC64556.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match      63.6%; Score 70; DB 2; Length 278;
```

Best Local Similarity 70.6%; Pred. No. 0.0072;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTGTGVLKQDWDGATI 18
: | | | | | | | | | | | | | | | |
Db 61 STVGVLKQDWDGSP 77

RESULT 24

ID Q9R8A6 PRELIMINARY; PRT; 278 AA.
AC Q9R8A6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DJ;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082748; AAC64554.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 63.6%; Score 70; DB 2; Length 278;

Best Local Similarity 70.6%; Pred. No. 0.0072;

Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTGTGVLKQDWDGATI 18
: | | | | | | | | | | | | | | | |
Db 61 STVGVLKQDWDGSP 77

RESULT 25

ID Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Louisiana;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082745; AAC64551.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 63.6%; Score 70; DB 2; Length 278;

Best Local Similarity 70.6%; Pred. No. 0.0072;

Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTGTGVLKQDWDGATI 18
: | | | | | | | | | | | | | | | |
Db 61 STVGVLKQDWDGSP 77

RESULT 26

ID Q9R8A5 PRELIMINARY; PRT; 278 AA.
AC Q9R8A5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fuzzy;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082749; AAC64555.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 63.6%; Score 70; DB 2; Length 278;

Best Local Similarity 70.6%; Pred. No. 0.0072;

Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTGTGVLKQDWDGATI 18
: | | | | | | | | | | | | | | | |
Db 61 STVGVLKQDWDGSP 77

RESULT 27

ID Q9ZGJ1 PRELIMINARY; PRT; 307 AA.
AC Q9ZGJ1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major outer membrane protein P30-1.
GN P30-1.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.

```
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL: AF078553; AAC68666.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 307 AA; 34108 MW; 9DA9FD63EBFBBC97 CRC64;

Query Match 63.6%; Score 70; DB 2; Length 307;
Best Local Similarity 70.6%; Pred. No. 0.008;
Matches 12; Conservative 2; Mismatches 0; Gaps 0;

QY 2 TTTGVFGLKQDWDGATI 18
Db 90 STVGVFGLKHDWDGSP 106

RESULT 28
Q8G8P3 PRELIMINARY; PRT; 291 AA.
ID Q8G8P3
AC Q8G8P3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479839; AAO12962.1; -.
DR EMBL: AF479840; AAO12968.1; -.
SQ SEQUENCE 291 AA; 32130 MW; 8220C436A1FC08E2 CRC64;

Query Match 60.0%; Score 66; DB 2; Length 291;
Best Local Similarity 61.1%; Pred. No. 0.032;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGATI 18
Db 60 NTTGVFGLKQDWDGSP 77

RESULT 29
Q8G948 PRELIMINARY; PRT; 276 AA.
ID Q8G948
AC Q8G948;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 15.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=West Paces, Heartland, St. Vincent, and Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479835; AAO12940.1; -.
DR EMBL: AF479836; AAO12945.1; -.
SQ SEQUENCE 276 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL: AF078553; AAC68666.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 307 AA; 34108 MW; 9DA9FD63EBFBBC97 CRC64;

Query Match 58.2%; Score 64; DB 2; Length 276;
Best Local Similarity 58.8%; Pred. No. 0.063;
Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGAT 17
Db 60 NATVALYGLKQDWNNGAS 76

RESULT 30
Q8GGU1 PRELIMINARY; PRT; 277 AA.
ID Q8GGU1
AC Q8GGU1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479838; AAO12957.1; -.
SQ SEQUENCE 277 AA; 30294 MW; 0CE7EDB51F2D854E CRC64;

Query Match 58.2%; Score 64; DB 2; Length 277;
Best Local Similarity 58.8%; Pred. No. 0.063;
Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGAT 17
Db 60 NATVAMYGLKQDWNNGAS 76

RESULT 31
Q8G8W7 PRELIMINARY; PRT; 277 AA.
ID Q8G8W7
AC Q8G8W7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479835; AAO12942.1; -.
DR EMBL: AF479836; AAO12947.1; -.
DR EMBL: AF479837; AAO12952.1; -.
SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

Query Match 58.2%; Score 64; DB 2; Length 277;
```

Best Local Similarity 58.8%; Pred. No. 0.063;
Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGAT 17
|||:|||||:|:
Db 60 NATVALYGLKQDWNQAS 76

RESULT 32

Q9AF98 PRELIMINARY; PRT; 270 AA.
AC Q9AF98;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE Major antigenic protein 1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cameroun;
RA Bensaïd A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF355203; AAK27219.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 270
SQ SEQUENCE 270 AA; 29380 MW; 99BAAFE9E48895A6 CRC64;

Query Match 55.9%; Score 61.5; DB 2; Length 270;
Best Local Similarity 66.7%; Pred. No. 0.15;
Matches 12; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 NTTTGVFLKQDWDGATI 18
|||:|||||:|:
Db 48 NTQT-VFGLKDKDWDGVQV 64

RESULT 33

Q93E65 PRELIMINARY; PRT; 277 AA.
AC Q93E65;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Blaukrans;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
Are Not Geographically Constrained and Show No Evidence of Having
Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL: AF368000; AAK98142.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 277
SQ SEQUENCE 277 AA; 29916 MW; 167B95113194237A CRC64;

Query Match 55.9%; Score 61.5; DB 2; Length 277;

Best Local Similarity 66.7%; Pred. No. 0.16;
Matches 12; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 NTTTGVFLKQDWDGATI 18
|||:|||||:|:
Db 55 NTQT-VFGLKDKDWDGVKV 71

RESULT 34

Q93E52 PRELIMINARY; PRT; 278 AA.
AC Q93E52;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=South East Botswana;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
Are Not Geographically Constrained and Show No Evidence of Having
Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL: AF368015; AAK98155.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 278
SQ SEQUENCE 278 AA; 30156 MW; 0355DE6B85842386 CRC64;

Query Match 55.9%; Score 61.5; DB 2; Length 278;
Best Local Similarity 66.7%; Pred. No. 0.16;
Matches 12; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 NTTTGVFLKQDWDGATI 18
|||:|||||:|:
Db 56 NTQT-VFGLKDKDWDGVK 72

RESULT 35

Q93E57 PRELIMINARY; PRT; 278 AA.
AC Q93E57;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Moringswaq2;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
Are Not Geographically Constrained and Show No Evidence of Having
Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203 (2001).
DR EMBL: AF368010; AAK98150.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 278

SQ SEQUENCE 278 AA; 30048 MW; 6C74669F156D4980 CRC64;
 Query Match 55.9%; Score 61.5; DB 2; Length 278;
 Best Local Similarity 66.7%; Pred. No. 0.16;
 Matches 12; Conservative 2; Mismatches 3; Indels 1; Gaps 1;
 QY 1 NTTTTGVFLKQDWDGATI 18
 ||| |||||:
 DB 56 NTQT-VFGLKKDWDGKV 72

Search completed: October 6, 2003, 07:52:31
 Job time : 30.1642 secs

83	42	39.3	131	22	AAU18909	Novel prostate gla
84	42	39.3	210	22	ABG22152	Novel human diagno
85	42	39.3	336	21	AGC28238	Arabidopsis thalia
86	42	39.3	330	22	AAH74663	C. elegans high af
87	42	39.3	576	22	ABH61695	Drosophila melanog
88	42	39.3	1002	22	ABH65804	Drosophila melanog
89	42	39.3	1002	22	ABH65811	Drosophila melanog
90	42	39.3	1002	22	ABH65834	Drosophila melanog
91	41	38.3	110	8	AAAP70092	Sequence encoded b
92	41	38.3	111	8	AAAP71705	Partial (2'-5') ol
93	41	38.3	111	16	AAH841187	Novel human diagno
94	41	38.3	131	22	ABG17311	Human (2'-5') olig
95	41	38.3	153	23	ABU51275	Helicobacter pylor
96	41	38.3	233	23	ABU52167	Helicobacter pylor
97	41	38.3	264	22	ABH60416	Drosophila melanog
98	41	38.3	272	23	ABU51878	Helicobacter pylor
99	41	38.3	282	22	AAU36009	Helicobacter pylor
100	41	38.3	364	8	AAAP70094	Sequence encoded b

ALIGNMENTS

RESULT 1
ABG30745
ID ABG30745 standard; Peptide; 20 AA.

21-OCT-2002 (first entry)

Ehrlichia chaffeensis peptide fragment #1.

KW Antibody detection; monoclonal antibody; polyclonal antibody.

OS Ehrlichia chaffeensis.

AA
PN WO200257794-A2.

AA
PD
25-JUL-2002.

16-JAN-2002; 2002WO-US01395.

XX
PR 18-JAN-2001: 2001US-0765739.XX
PA (INDEX-) IDEXX LAB INC.

XX
PI Lawton R. O'Connor TP. Barto] RA. Machenry ps.

XX
DR WPT: 2002-599730/64

XX
PT New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to *Ehrlichia canis* or *chaffeensis*, or in
PT detecting or quantifying the presence of *Ehrlichia* infection in mammals

PS Claim 1; Page 5; 29pp; English.

The invention relates to a composition of matter comprising a polypeptide isolated from *Ehrlichia* species. The composition can be used for detecting the presence of antibodies to *Ehrlichia*, comprising contacting one or more polypeptides with a test sample suspected of comprising antibodies to *Ehrlichia*, under conditions that allow polypeptide/antibody complexes to form and detecting the complexes, where the detection of polypeptide/antibody complexes is an indication that antibodies to *Ehrlichia* are present in the test sample. The composition is useful for detecting or quantifying the presence of *E. canis* or *E. chaffeensis* infection in mammals. The polypeptides can be used to develop monoclonal and/or polyclonal antibodies that can be employed in assay systems and in the generation of chimeric antibodies for therapeutic use or other similar applications. This sequence represents an *E. chaffeensis* peptide fragment used in the composition of the invention.

SQ	Sequence	20 AA;
Query Match	100.0%;	Score 107; DB 23; Length 20;
Best Local Similarity	100.0%;	Pred. No. 2.8e-10;
Matches 20;	Conservative	0; Mismatches 0; Indels 0; Caps 0;

RESULT 2
 AAY06942
 ID AAY06942 standard; Protein; 256 AA.

DT 05-JUL-1999 (first entry)

DE *E. chafeensis* p28 protein.

Outer membrane protein; OMP; *Ehrlichia chafeensis*; *E. canis*; P30; detection; dog. KW

OS Ehrlichia chafeensis.

PN WO9913720-A1.

25-MAR-1999.

AA 18-SEP-1998; 98WO-US19600.
PF

XX	19-SEP-1997;	97US-0059353.
PR		

XX
PA (OHIS) UNIV OHIO STATE.

XX Ohashi N. Rikihisa Y:

DR WPI; 1999-254290/21.

DR N-PSDB; AAX34742.

PT	Novel outer membrane proteins from <i>Ehrlichia chaffeensis</i> and <i>Ehrlichia canis</i>
PT	
PT	

PS Claim 18; Fig 1; 55pp; English.

The invention provides isolated outer membrane proteins (OMP) from *Escherichia chaffeensis* and *E. canis*. The *E. chaffeensis* proteins form part of the OMP family and consist of proteins OMP-1, -1(B to z) shown in *RAY05943-958*. The *E. canis* proteins form part of the P30 family and consist of proteins shown in *RAY06959-970*. The proteins and genes are used to detect *E. chaffeensis* in patients and *E. canis* in dogs.

AX	Sequence	256 AA;
----	----------	---------

RESULT 3
ABG77966
ID ABG77966 standard; Protein; 256 AA.

DT	15-NOV-2002	(first entry)
----	-------------	---------------

XX DE Protein encoded by *Ehrlichia chaffeensis* p28 gene.

```

XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection; p28; OMP-1.
XX OS Ehrlichia chaffeensis.
XX PN US2002120115-A1.
XX PD 29-AUG-2002.
XX PF 28-JAN-2002; 2002US-0059964.
XX PR 19-MAY-1999; 99US-0314701.
XX PA (RIKI/) RIKIHISA Y.
XX PA (OHAS/) OHASHI N.
XX PI Rikihisa Y, Ohashi N;
XX WPI; 2002-618954/66.
XX DR N-PSDB; ABS63307.
XX PT Isolated polynucleotide encoding an outer membrane protein of E.canis
XX PT or E.chaffeensis used in the diagnosis of infection -
XX PS Disclosure; Fig 1; 49pp; English.
XX CC The invention relates to an isolated polynucleotide encoding an outer
XX CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX CC in the diagnosis of infection. An infection such as human ehrlichiosis or
XX CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX CC patient, providing a polypeptide or mixture of polypeptides, contacting
XX CC the sample with the polypeptide and assaying for the formation of a
XX CC complex between antibodies in the serum sample and the polypeptide, where
XX CC formation of a complex is indicative of infection with E. chaffeensis.
XX CC This sequence represents the Ehrlichia chaffeensis OMP-1 protein encoded
XX CC by the p28 gene.
XX SQ Sequence 256 AA;

Query Match 100.0%; Score 107; DB 23; Length 256;
Best Local Similarity 100.0%; Pred. No. 6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
Db 34 NTTGVFGLKQNWGSAISN 53

RESULT 4
AAW51095
ID AAW51095 standard; Protein; 276 AA.
XX AC AAW51095;
XX DT 14-SEP-1998 (first entry)
XX DE Ehrlichia chaffeensis VSA5 protein (partial sequence).
XX KW MAP1 homologue; variable surface antigen; VSA5; rickettsia;
XX KW DNA vaccine.
XX OS Ehrlichia chaffeensis.
XX FH Key Location/Qualifiers
XX FT Peptide 1..25
XX FT /note= "putative signal peptide"
XX PN WO9816554-A1.
XX PD 23-APR-1998.
XX PF 17-OCT-1997; 97WO-US19044.
XX

```

```

PR 17-OCT-1996; 96US-0733230.
XX (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Burr ridge MJ, Ganta RR, Mahan SM, McGuire TC;
XX PI Nyika A, Rurangirwa FR;
XX WPI; 1998-251232/22.
XX DR N-PSDB; AAV07179.
XX PT Composition containing nucleic acid encoding rickettsial antigen -
XX PT useful for, e.g. stimulating protective immune response in humans or
XX PT animals
XX PS Claim 3; Fig 2B; 39pp; English.
XX CC This is the near full-length variable surface antigen VSA5 protein
XX CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
XX CC residues. The VSA5 amino acid sequence was deduced from a partial
XX CC open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
XX CC chaffeensis that was obtained on the basis of homology to the major
XX CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
XX CC genomic locus included 5 ORFs encoding similar, but non-identical
XX CC proteins (see AAW51091-95). A claimed composition comprises a
XX CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
XX CC that elicits a protective immune response against a rickettsial
XX CC pathogen. The nucleic acid is used, in human or veterinary
XX CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,
XX CC Anaplasma and Cowdria species. The Ehrlichia antigenic
XX CC polypeptides can be used diagnostically to detect antibodies
XX CC associated with Ehrlichia infection (claimed).
XX SQ Sequence 276 AA;

Query Match 100.0%; Score 107; DB 19; Length 276;
Best Local Similarity 100.0%; Pred. No. 6.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
Db 59 NTTGVFGLKQNWGSAISN 78

RESULT 5
AAB36189
ID AAB36189 standard; Protein; 276 AA.
XX AC AAB36189;
XX DT 02-MAR-2001 (first entry)
XX DE Ehrlichia chaffeensis partial VSA5.
XX KW Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
XX KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
XX KW Rickettsia; Cowdria; Anaplasma; map2; 1hworf3; 4hworf1; 18hworf1;
XX KW 3gdorf3.
XX OS Ehrlichia chaffeensis.
XX PN WO2000065063-A2.
XX PD 02-NOV-2000.
XX PF 21-APR-2000; 2000WO-US10886.
XX PR 22-APR-1999; 99US-0130725.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Bowie MV, Ganta RR, Burr ridge MJ, Mahan SM, McGuire TC;
XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX

```



```
SQ Sequence 280 AA;
Query Match 100.0%; Score 107; DB 19; Length 280;
Best Local Similarity 100.0%; Pred. No. 6.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
   |||||
DB 60 NTTGVFGLKQNWGSAISN 79

RESULT 8
AAB36183
ID AAB36183 standard; Protein; 280 AA.
XX
AC AAB36183;
XX
DT 02-MAR-2001 (first entry)
XX
DE Ehrlichia chaffeensis MAP1.
XX
KW Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
KW 4hworf1; 18hworf1; 3gdoorf3.
XX
XX Ehrlichia chaffeensis.
OS
XX WO200065063-A2.
PN
XX
XX 02-NOV-2000.
PD
XX
XX 21-APR-2000; 2000WO-US10886.
PF
XX
XX 22-APR-1999; 99US-0130725.
PR
XX
XX (UYFL ) UNIV FLORIDA.
PA
XX
XX Barbet AF, Bowie MV, Ganta RR, Burrridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX
XX WPI; 2000-679675/66.
DR N-PSDB; AAC68700.
XX
XX New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
PT
XX
XX Claim 3; Page 35-36; 63pp; English.
PS
XX
XX The present sequence is given in a specification relating to nucleic
CC acid vaccines containing genes to protect animals or humans against
CC rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
CC sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The vaccine comprises the
CC major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
CC (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
CC driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
CC ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1 and
CC 3gdoorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.
XX
SQ Sequence 280 AA;
Query Match 100.0%; Score 107; DB 21; Length 280;
Best Local Similarity 100.0%; Pred. No. 6.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
   |||||
DB 60 NTTGVFGLKQNWGSAISN 79

RESULT 9
AAU04193
ID AAU04193 standard; Protein; 280 AA.
XX
AC AAU04193;
XX
DT 23-OCT-2001 (first entry)
XX
DE Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.
XX
KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic.
XX
XX Ehrlichia chaffeensis.
OS
XX US6251872-B1.
PN
XX
XX 26-JUN-2001.
PD
XX
XX 17-OCT-1997; 97US-0953326.
PF
XX
XX 17-OCT-1996; 96US-0733230.
PR
XX
XX (UYFL ) UNIV FLORIDA.
PA
XX
XX Barbet AF, Ganta RR, McGuire TC, Burrridge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
XX WPI; 2001-424487/45.
DR N-PSDB; AAS07576.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures
PT
XX
XX Disclosure; Column 15-17; 30pp; English.
PS
XX
XX The sequence represents the amino acid sequence of major antigenic
CC protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
CC polypeptides are useful as vaccines for conferring immunity to rickettsia
CC infection, including Cowdria ruminantium causing heartwater. The MAP
CC polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.
XX
SQ Sequence 280 AA;
Query Match 100.0%; Score 107; DB 22; Length 280;
Best Local Similarity 100.0%; Pred. No. 6.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
   |||||
DB 60 NTTGVFGLKQNWGSAISN 79

RESULT 10
AAY06943
ID AAY06943 standard; Protein; 281 AA.
XX
AC AAY06943;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. chaffeensis OMP-1 protein.
```

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO9913720-A1.
 XX
 PD 25-MAR-1999.
 XX
 XX 18-SEP-1998; 98WO-US19600.
 PF
 XX 19-SEP-1997; 97US-0059353.
 PR
 XX (OHIS) UNIV OHIO STATE.
 PA
 PI Ohashi N, Rikihisa Y;
 XX
 XX WPI; 1999-254290/21.
 DR
 DR N-PSDB; AAX34743.
 XX
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 PT
 PS Disclosure; Fig 3B; 55pp; English.
 XX
 CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1-11(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 CC
 XX
 SQ Sequence 281 AA;

 Query Match 100.0%; Score 107; DB 20; Length 281;
 Best Local Similarity 100.0%; Pred. No. 6.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 NTTGVFGLKQNDGSAISN 20
 Db 59 NTTGVFGLKQNDGSAISN 78

 RESULT 11
 ABG77935
 ID ABG77935 standard; Protein; 281 AA.
 XX
 AC ABG77935;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Ehrlichia chaffeensis outer membrane protein (OMP) #1.
 XX
 KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN US2002120115-A1.
 XX
 PD 29-AUG-2002.
 XX
 PF 28-JAN-2002; 2002US-0059964.
 XX
 PR 19-MAY-1999; 99US-0314701.
 XX
 PA (RIKI/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.
 XX
 PI Rikihisa Y, Ohashi N;
 XX
 XX WPI; 2002-618954/66.
 DR
 DR N-PSDB; ABS63276.

XX Isolated polynucleotide encoding an outer membrane protein of E. canis
 PT or E. chaffeensis used in the diagnosis of infection -
 XX
 XX Claim 14; Fig 3B; 49pp; English.
 PS
 XX
 CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.
 XX
 SQ Sequence 281 AA;

 Query Match 100.0%; Score 107; DB 23; Length 281;
 Best Local Similarity 100.0%; Pred. No. 6.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Qy 1 NTTGVFGLKQNDGSAISN 20
 Db 59 NTTGVFGLKQNDGSAISN 78

 RESULT 12
 AAU96105
 ID AAU96105 standard; Protein; 281 AA.
 XX
 AC AAU96105;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Ehrlichia chaffeensis P28.
 XX
 KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
 KW antibacterial.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO200222782-A2.
 XX
 PD 21-MAR-2002.
 XX
 PF 12-SEP-2001; 2001WO-US28759.
 XX
 PR 12-SEP-2000; 2000US-0660587.
 XX
 PA (RERE-) RES DEV FOUND.
 XX
 PI Walker DH, Yu X, McBride JW;
 XX
 DR WPI; 2002-351882/38.
 XX
 PT New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
 XX
 PS Example 3; Figure 3; 106pp; English.
 XX
 CC The invention relates to a recombinant homologous 28 kDa immunodominant
 CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX
 SQ Sequence 281 AA;

 Query Match 100.0%; Score 107; DB 23; Length 281;

Best Local Similarity 100.0%; Pred. No. 6.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAISN 20
Db 59 NTTGVGFLKQKQNDGSAISN 78

RESULT 13

AAU73418
ID AAU73418 standard; Protein; 281 AA.

XX AC

XX AAU73418;

XX 12-MAR-2002 (first entry)

XX Ehrlichia chaffeensis outer membrane protein P28-19.

XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.

XX Ehrlichia chaffeensis.

XX WO200183699-A2.

XX 08-NOV-2001.

XX 01-MAY-2001; 2001WO-US13997.

XX 01-MAY-2000; 2000US-201035P.

XX (RERE-) RES DEV FOUND.

XX Walker DH, Yu X;

XX WPI; 2002-066527/09.

XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated

PT P28 useful as a vaccine against Ehrlichia chaffeensis

PT Disclosure; Figure 2; 97pp; English.

XX The invention relates to isolated and purified 28-kDa outer membrane

CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins

CC are encoded by a 28kDa outer membrane protein multigene family. P28

CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28

CC is useful for transfecting a host cell. AAU73400-AAU73420 represent

CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.

XX SQ Sequence 281 AA;

Query Match

Best Local Similarity 100.0%; Score 107; DB 23; Length 281;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAISN 20

Db 59 NTTGVGFLKQKQNDGSAISN 78

RESULT 14

AAW51094
ID AAW51094 standard; Protein; 280 AA.

XX AC

XX AAW51094;

XX 14-SEP-1998 (first entry)

XX Ehrlichia chaffeensis VSA4 protein.

XX MAP1 homologue; variable surface antigen; VSA4; rickettsia;

XX DNA vaccine.

XX Ehrlichia chaffeensis.

XX Key Location/Qualifiers
FH Peptide 1..25
FT /note= "putative signal peptide"

XX WO9816554-A1.

XX 23-APR-1998.

XX 17-OCT-1997; 97WO-US19044.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Burrige MJ, Ganta RR, Mahan SM, McGuire TC;

PI Nyika A, Rurangirwa FR;

XX WPI; 1998-251232/22.

XX N-PSDB; AAV07179.

XX Composition containing nucleic acid encoding rickettsial antigen -
PT useful for, e.g. stimulating protective immune response in humans or
PT animals

PS Claim 3; Fig 2B; 39pp; English.

XX This is the full-length variable surface antigen VSA4 protein of
CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)
CC of E. chaffeensis that was obtained on the basis of homology to the
CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
CC This genomic locus included 5 ORFs encoding similar, but
CC non-identical proteins (see AAW51091-95). A claimed composition
CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
CC (see AAW51088-99) that elicits a protective immune response against a
CC rickettsial pathogen. The nucleic acid is used, in human or
CC veterinary medicine, in vaccines to protect against Rickettsia,
CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
CC polypeptides can be used diagnostically to detect antibodies
CC associated with Ehrlichia infection (claimed).

XX Sequence 280 AA;

Query Match 85.0%; Score 91; DB 19; Length 280;

Best Local Similarity 84.2%; Pred. No. 2.5e-06;

Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAIS 19

Db 60 NTTGVGFLKQKQNDGSAIS 78

RESULT 15

AAB36188

ID AAB36188 standard; Protein; 280 AA.

XX AC

XX AAB36188;

XX 02-MAR-2001 (first entry)

XX Ehrlichia chaffeensis partial VSA4.

XX Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;

XX major antigenic protein 1; antirickettsial; vaccine; gene therapy;

XX Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;

XX 3gdorf3.

XX Ehrlichia chaffeensis.

XX WO200065063-A2.

XX 02-NOV-2000.

```
XX PF 21-APR-2000; 2000WO-US10886.
XX PR 22-APR-1999; 99US-0130725.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX DR N-PSDB; AAC68705.
XX DR 2000-679675/66.
XX PT New polynucleotides useful as DNA vaccines for conferring immunity to
XX PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX PS Claim 3; Page 45-46; 63pp; English.
XX CC The present sequence shows a high degree of similarity to the major
XX CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
XX CC used in a vaccines to protect animals or humans against rickettsial
XX CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
XX CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
XX CC protective against the rickettsial pathogen. The nucleic acid vaccines
XX CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
XX CC Cowdria ruminantium genes designated map 2, lhwr3, 4hwr3, 18hwr3,
XX CC and 3gdr3 may be used in therapeutic and diagnostic applications. The
XX CC polypeptides are useful for detecting antibodies associated with
XX CC infection by a rickettsial pathogen whilst the polynucleotides may be
XX CC used to detect the presence of rickettsial nucleic acids.
XX SQ Sequence 280 AA;
Query Match 85.0%; Score 91; DB 21; Length 280;
Best Local Similarity 84.2%; Pred. No. 2.5e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQNDGSAIS 19
DB |||:|||||:|||||
60 NTTIGVFLKQNDGSGTIS 78
RESULT 16
AAU04198
XX ID AAU04198 standard; Protein; 280 AA.
XX AC AAU04198;
XX DT 23-OCT-2001 (first entry)
XX DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
XX KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
XX KW infection; heartwater; diagnostic; variable surface antigen; VSA.
XX OS Ehrlichia chaffeensis.
XX PN US6251872-B1.
XX PD 26-JUN-2001.
XX PF 17-OCT-1997; 97US-0953326.
XX PR 17-OCT-1996; 96US-0733230.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX DR WPI; 2001-424487/45.
DR N-PSDB; AAS07578.
XX PT New MAP2 genes and polypeptides useful as vaccines for conferring
XX PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX PT as molecular markers in nucleic acid analysis procedures -
XX PS Example 3; Fig 2A-2B; 30pp; English.
XX CC The sequence represents the amino acid sequence of variable surface
XX CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
XX CC has similarity to major antigen protein (MAP). The MAP polynucleotides
XX CC and polypeptides are useful as vaccines for conferring immunity to
XX CC rickettsial infection, including Cowdria ruminantium causing heartwater.
XX CC The MAP polynucleotides may be used as molecular markers in nucleic acid
XX CC analysis procedures, and to produce the MAP polypeptides, which may
XX CC be used to raise antibodies that are reactive with the polypeptides.
XX CC The nucleic acids may further be used as probes to identify
XX CC complementary sequences within other nucleic acid molecules or genomes,
XX CC where such probes can be applied to identify or distinguish infectious
XX CC strains of organisms in diagnostic procedures or in rickettsial
XX CC research where identification of particular organisms or strains is
XX CC needed.
XX SQ Sequence 280 AA;
Query Match 85.0%; Score 91; DB 22; Length 280;
Best Local Similarity 84.2%; Pred. No. 2.5e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQNDGSAIS 19
DB |||:|||||:|||||
60 NTTIGVFLKQNDGSGTIS 78
RESULT 17
AAU73417
XX ID AAU73417 standard; Protein; 280 AA.
XX AC AAU73417;
XX DT 12-MAR-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein P28-18.
XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX OS Ehrlichia chaffeensis.
XX PN WO200183699-A2.
XX PD 08-NOV-2001.
XX PF 01-MAY-2001; 2001WO-US13997.
XX PR 01-MAY-2000; 2000US-201035P.
XX PA (BERE-) RES DEV FOUND.
XX PI Walker DH, Yu X;
XX DR WPI; 2002-066527/09.
XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
XX PT P28 useful as a vaccine against Ehrlichia chaffeensis -
XX PS Disclosure; Figure 2; 97pp; English.
XX CC The invention relates to isolated and purified 28-kDa outer membrane
XX CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
XX CC are encoded by a 28kDa outer membrane protein multigene family. P28
XX CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
XX CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
XX CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
```

```
XX SQ Sequence 280 AA;
Query Match 85.0%; Score 91; DB 23; Length 280;
Best Local Similarity 84.2%; Pred. No. 2.5e-06;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
   |||:|||||:|||| ||
Db 60 NTTGVFGLKQDWDGSTIS 78

RESULT 18
ABG30749
ID ABG30749 standard; Peptide; 19 AA.
XX AC ABG30749;
XX DT 21-OCT-2002 (first entry)
XX DE Ehrlichia chaffeensis peptide fragment #5.
XX KW Antibody detection; monoclonal antibody; polyclonal antibody.
XX OS Ehrlichia chaffeensis.
XX PN WO200257794-A2.
XX PD 25-JUL-2002.
XX PF 16-JAN-2002; 2002WO-US01395.
XX PR 18-JAN-2001; 2001US-0765739.
XX PA (INDEX-) IDEXX LAB INC.
XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX DR WPI; 2002-599730/64.
XX PT New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
PT
XX Claim 1; Page 5; 29pp; English.
XX CC The invention relates to a composition of matter comprising a polypeptide
XX isolated from Ehrlichia species. The composition can be used for
XX detecting the presence of antibodies to Ehrlichia, comprising contacting
XX one or more polypeptides with a test sample suspected of comprising
XX antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
XX complexes to form and detecting the complexes, where the detection of
XX polypeptide/antibody complexes is an indication that antibodies to
XX Ehrlichia are present in the test sample. The composition is useful for
XX detecting or quantifying the presence of E. canis or E. chaffeensis
XX infection in mammals. The polypeptides can be used to develop monoclonal
XX and/or polyclonal antibodies that can be employed in assay systems and in
XX the generation of chimeric antibodies for therapeutic use or other
XX similar applications. This sequence represents an E. chaffeensis peptide
XX fragment used in the composition of the invention.
XX SQ Sequence 19 AA;

Query Match 82.2%; Score 88; DB 23; Length 19;
Best Local Similarity 84.2%; Pred. No. 3e-07;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
   |||:|||||:|||| ||
Db 1 NTTGVFGLKQDWDGSTIS 19

RESULT 19
AAAY06948
ID AAY06948 standard; Protein; 280 AA.
XX AC AAY06948;
XX DT 05-JUL-1999 (first entry)
XX DE E. chaffeensis OMP-1F protein.
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.
XX OS Ehrlichia chaffeensis.
XX PN WO9913720-A1.
XX PD 25-MAR-1999.
XX PF 18-SEP-1998; 98WO-US19600.
XX PR 19-SEP-1997; 97US-0059353.
XX PA (OHIS ) UNIV OHIO STATE.
XX PI Ohashi N, Rikihisa Y;
XX DR WPI; 1999-254290/21.
XX DR N-PSDB; AAX34748.
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX Claim 16; Fig 8B; 55pp; English.
XX CC The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX SQ Sequence 280 AA;

Query Match 82.2%; Score 88; DB 20; Length 280;
Best Local Similarity 84.2%; Pred. No. 7.5e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
   |||:|||||:|||| ||
Db 60 NTTGVFGLKQDWDGSTIS 78

RESULT 20
ABG77940
ID ABG77940 standard; Protein; 280 AA.
XX AC ABG77940;
XX DT 15-NOV-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein (OMP) #6.
XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX OS Ehrlichia chaffeensis.
XX PN US2002120115-A1.
XX PD 29-AUG-2002.
XX PR 28-JAN-2002; 2002US-0059964.
XX
```

PR 19-MAY-1999; 99US-0314701.
XX (RKI/) RIKIHISA Y.
PA (OHAS/) OHASHI N.
XX Rikihisa Y, Ohashi N;
XX WPI; 2002-618954/66.
DR N-PSDB; ABS63281.
XX Isolated polynucleotide encoding an outer membrane protein of E.canis
PT or E.chaffeensis used in the diagnosis of infection -
XX Disclosure; Fig 8B; 49pp; English.
XX The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX Sequence 280 AA;
SQ
Query Match 82.2%; Score 88; DB 23; Length 280;
Best Local Similarity 84.2%; Pred. No. 7.5e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSTIS 78
RESULT 21.
AAU96110
ID AAU96110 standard; Protein; 280 AA.
XX AAU96110;
AC
DT 02-JUL-2002 (first entry)
XX Ehrlichia chaffeensis OMP-1F.
DE Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
KW antibacterial.
XX Ehrlichia chaffeensis.
OS
PN WO200222782-A2.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28759.
XX
PR 12-SEP-2000; 2000US-0660587.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X, McBride JW;
XX WPI; 2002-351882/38.
DR
XX New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX
PS Example 3; Figure 3; 106pp; English.
XX The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably

CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
SQ Sequence 280 AA;
Query Match 82.2%; Score 88; DB 23; Length 280;
Best Local Similarity 84.2%; Pred. No. 7.5e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSTIS 78
RESULT 22
ABG30744
ID ABG30744 standard; Peptide; 20 AA.
XX
AC ABG30744;
XX
DT 21-OCT-2002 (first entry)
XX Ehrlichia canis peptide fragment #2.
DE
XX Antibody detection; monoclonal antibody; polyclonal antibody.
KW Ehrlichia canis.
XX
OS
PN WO200257794-A2.
XX
PD 25-JUL-2002.
XX
PF 16-JAN-2002; 2002WO-US01395.
XX
PR 18-JAN-2001; 2001US-0765739.
XX
PA (IDEX-) IDEXX LAB INC.
XX
PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX WPI; 2002-599730/64.
DR
XX
PS New composition of matter comprising a polypeptide, useful in detecting
CC the presence of antibodies to Ehrlichia canis or chaffeensis, or in
CC detecting or quantifying the presence of Ehrlichia infection in mammals
CC
XX Claim 1; Page 5; 29pp; English.
XX The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. canis peptide
CC fragment used in the composition of the invention.
XX
SQ Sequence 20 AA;
Query Match 76.6%; Score 82; DB 23; Length 20;
Best Local Similarity 70.0%; Pred. No. 2.9e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAISN 20
 DB 1 NTTGVGFLKQKQNDGATIKD 20

RESULT 23

AA06959
 ID AAY06959 standard; Protein; 288 AA.

AC AAY06959;

XX 05-JUL-1999 (first entry)

DT E. canis P30 protein.

DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

XX Ehrlichia canis.

XX WO9913720-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

XX 19-SEP-1997; 97US-0059353.

XX (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI; 1999-254290/21.

DR N-PSDB; AAX34759.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis

PS Disclosure; Fig 19B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 288 AA;

Query Match 76.6%; Score 82; DB 20; Length 288;
 Best Local Similarity 70.0%; Pred. No. 7.1e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAISN 20
 DB 60 NTTGVGFLKQKQNDGATIKD 79

RESULT 24

ABG77950
 ID ABG77950 standard; Protein; 288 AA.

XX ABG77950;

XX 15-NOV-2002 (first entry)

XX Ehrlichia canis outer membrane protein (P30F) #1.

XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

XX Ehrlichia canis.

XX US2002120115-A1.

PN

XX 29-AUG-2002.
 PD 28-JAN-2002; 2002US-0059964.
 XX PF 19-MAY-1999; 99US-0314701.

XX (RIKI/) RIKIHISA Y.

PA (OHAS/) OHASHI N.

XX Rikihisa Y, Ohashi N;

XX WPI; 2002-618954/66.

XX N-PSDB; ABS63291.

XX Isolated polynucleotide encoding an outer membrane protein of E. canis
 PT or E. chaffeensis used in the diagnosis of infection -
 PT

XX Claim 10; Fig 19B; 49pp; English.

XX The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.

XX Sequence 288 AA;

Query Match 76.6%; Score 82; DB 23; Length 288;
 Best Local Similarity 70.0%; Pred. No. 7.1e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAISN 20
 DB 60 NTTGVGFLKQKQNDGATIKD 79

RESULT 25

AA06962

ID AAY06962 standard; Protein; 280 AA.

XX AAY06962;

XX 05-JUL-1999 (first entry)

XX E. canis P30-2 protein.

XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

XX Ehrlichia canis.

XX WO9913720-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

XX 19-SEP-1997; 97US-0059353.

XX (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI; 1999-254290/21.

XX N-PSDB; AAX34762.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT

PT Ehrlichia canis
 PS Disclosure; Fig 22B; 55pp; English.
 XX
 CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AA06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AA06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 SQ Sequence 280 AA;
 Query Match 74.8%; Score 80; DB 20; Length 280;
 Best Local Similarity 70.0%; Pred. No. 0.00014;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVFGLKQNWGSAISN 20
 I:|||||||:|:|
 DB 59 NSTGVFGLKHDWNGGTISN 78
 I:|||||||:|:|
 RESULT 26
 AA071479
 ID AA071479 standard; Protein; 280 AA.
 XX
 AC AA071479;
 XX
 DT 12-OCT-2000 (first entry)
 XX
 DE Ehrlichia canis immunoreactive protein ECA28SA3.
 XX
 KW Homologous 28-kDa protein gene; ECA28SA3; immunoreactive; vaccine;
 KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
 KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
 KW tick-borne rickettsial disease; serodiagnosis.
 XX
 OS Ehrlichia canis.
 XX
 EH Key Location/Qualifiers
 FH Peptide 1..23
 FT /label= Signal_peptide
 FT Protein 24..280
 FT /label= Mature_ECA28SA3_28-kDa_protein
 XX
 PN WO200032745-A2.
 XX
 PD 08-JUN-2000.
 XX
 PF 24-NOV-1999; 99WO-US28075.
 XX
 PR 30-NOV-1998; 98US-0201458.
 PR 03-MAR-1999; 99US-0261358.
 XX
 PA (RERE-) RES DEV FOUND.
 XX
 PI Walker DH, Yu X, McBride JW;
 XX WPI; 2000-412298/35.
 DR N-PSDB; AAD01294, AAD01295.
 XX
 PT Ehrlichia canis antigens useful for vaccinating against canine
 PT ehrlichiosis in dogs -
 XX
 PS Claim 12; Page 68-69; 86pp; English.
 XX
 CC The patent relates to homologous 28-kilodalton (kDa) protein genes of
 CC Ehrlichia canis, designated ECA28SA1, ECA28SA2, ECA28SA3, ECA28-1 and
 CC ECA28-2. These genes are members of a polymorphic multiple gene family
 CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are
 CC immunoreactive with anti-E. canis serum hence are important
 CC immunoprotective antigens. The protein is useful for vaccinating
 CC against E. canis infections such as canine ehrlichiosis in dogs.

CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
 CC tick-borne rickettsial disease of dogs. ECA28-1 is conserved amongst
 CC different strains of E. canis and hence useful for serodiagnosis of
 CC canine ehrlichiosis. The present sequence is a E. canis
 CC ECA28SA3 30-kDa protein which is post-translationally modified to a
 CC mature 28-kDa protein by cleavage of N-terminal signal sequence.
 XX
 SQ Sequence 280 AA;
 Query Match 74.8%; Score 80; DB 21; Length 280;
 Best Local Similarity 70.0%; Pred. No. 0.00014;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVFGLKQNWGSAISN 20
 I:|||||||:|:|
 DB 59 NSTGVFGLKHDWNGGTISN 78
 I:|||||||:|:|
 RESULT 27
 ABG77953
 ID ABG77953 standard; Protein; 280 AA.
 XX
 AC ABG77953;
 XX
 DT 15-NOV-2002 (first entry)
 XX
 DE Ehrlichia canis outer membrane protein (P30F) #4.
 XX
 KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
 XX
 OS Ehrlichia canis.
 XX
 PN US2002120115-A1.
 XX
 PD 29-AUG-2002.
 XX
 PF 28-JAN-2002; 2002US-0059964.
 XX
 PR 19-MAY-1999; 99US-0314701.
 PA (RIKI/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.
 XX
 PI Rikihisa Y, Ohashi N;
 XX WPI; 2002-618954/66.
 DR N-PSDB; ABS63294.
 XX
 PT Isolated polynucleotide encoding an outer membrane protein of E. canis
 PT or E. chaffeensis used in the diagnosis of infection -
 XX
 PS Claim 10; Fig 22B; 49pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.
 XX
 SQ Sequence 280 AA;
 Query Match 74.8%; Score 80; DB 23; Length 280;
 Best Local Similarity 70.0%; Pred. No. 0.00014;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVFGLKQNWGSAISN 20
 I:|||||||:|:|
 DB 59 NSTGVFGLKHDWNGGTISN 78
 I:|||||||:|:|

	XX
	Pf
16-JAN-2002; 2002WO-US01395.	CC
18-JAN-2001; 2001US-0765739.	CC
(IDEX-) IDEXX LAB INC.	PR
Lawton R, O'Connor TP, Bartol BA, Machenry PS;	XX
WPI; 2002-599730/64.	PA
New composition of matter comprising a polypeptide, useful in detecting the presence of antibodies to Ehrlichia canis or chaffeensis, or in detecting or quantifying the presence of Ehrlichia infection in mammals	XX
.	PT
	XX
Claim 1; Page 5; 29pp; English.	PS
The invention relates to a composition of matter comprising a polypeptide isolated from Ehrlichia species. The composition can be used for detecting the presence of antibodies to Ehrlichia, comprising contacting one or more polypeptides with a test sample suspected of comprising antibodies to Ehrlichia, under conditions that allow polypeptide/antibody complexes to form and detecting the complexes, where the detection of polypeptide/antibody complexes is an indication that antibodies to Ehrlichia are present in the test sample. The composition is useful for detecting or quantifying the presence of E. canis or E. chaffeensis infection in mammals. The polypeptides can be used to develop monoclonal and/or polyclonal antibodies that can be employed in assay systems and in the generation of chimeric antibodies for therapeutic use or other similar applications. This sequence represents an E. chaffeensis peptide fragment used in the composition of the invention.	XX
	XX
Sequence 19 AA;	SQ
Query Match 68.2%; Score 73; DB 23; Length 19;	
Best Local Similarity 68.4%; Pred. No. 7.5e+05;	
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;	
QY 1 NTTVGVLGKQNWDGSNIS 19	
Dd :: ::	
Bb 1 NTTVGVGEIQDWRDVIS 19	
RESULT 30	
AAW51092	
ID AAW51092 standard; Protein; 286 AA.	
XX AC AAW51092:	
XX DT	
DT 14-SEP-1998 (first entry)	
XX DE Ehrlichia chaffeensis VSA2 protein.	
DE KW MAP1 homologue; variable surface antigen; VSA2; rickettsia; DNA vaccine.	
KX OS Ehrlichia chaffeensis.	
OS XX Key Location/Qualifiers	
FH FT Peptide 1..25 /note= "putative signal peptide"	
FT FT	
WW W09816554-AI.	
PN PD 23-APR-1998.	
PD XX 17-OCT-1997; 97WO-US19044.	
XX PF 17-OCT-1996; 96US-0733230.	
PF PR (UYFL) UNIV FLORIDA.	
PR PA	
PA XX	

PI Barbet AF, Burr ridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX WPI: 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX Claim 3; Fig 2A; 39pp; English.
 PS This is the full-length variable surface antigen VSA2 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF2) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX Sequence 286 AA;
 SQ
 Query Match 68.2%; Score 73; DB 19; Length 286;
 Best Local Similarity 68.4%; Pred. No. 0.0019;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVGFLKQNWGSAIS 19
 DB 60 NTTGVGFLKQNWGSAIS 78
 RESULT 31
 AAY06946
 ID AAY06946 standard; Protein: 286 AA.
 XX
 AC AAY06946;
 DT 05-JUL-1999 (first entry)
 DE E. chaffeensis OMP-1D protein.
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX Ehrlichia chaffeensis.
 OS WO9913720-A1.
 XX
 PN 25-MAR-1999.
 PD 18-SEP-1998; 98WO-US19600.
 PF 19-SEP-1997; 97US-0059353.
 PR (OHIS) UNIV OHIO STATE.
 PA Ohashi N, Rikihisa Y;
 PI WPI: 1999-254290/21.
 DR N-PSDB; AAX34746.
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 PT Claim 14; Fig 6B; 55pp; English.
 PS
 XX

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX Sequence 286 AA;
 SQ
 Query Match 68.2%; Score 73; DB 20; Length 286;
 Best Local Similarity 68.4%; Pred. No. 0.0019;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVGFLKQNWGSAIS 19
 DB 60 NTTGVGFLKQNWGSAIS 78
 RESULT 32
 AAB36186
 ID AAB36186 standard; Protein: 286 AA.
 XX
 AC AAB36186;
 DT 02-MAR-2001 (first entry)
 DE Ehrlichia chaffeensis partial VSA2.
 XX
 KW Ehrlichia chaffeensis; VSA2; variable surface antigen 2; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX Ehrlichia chaffeensis.
 OS WO200065063-A2.
 XX
 PN 02-NOV-2000.
 PD 21-APR-2000; 2000WO-US10886.
 PF 22-APR-1999; 99US-0130725.
 PR (UYFL) UNIV FLORIDA.
 PA Barbet AF, Bowie MV, Ganta RR, Burr ridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 XX WPI: 2000-679675/66.
 DR N-PSDB; AAC68703.
 XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 PT
 XX Claim 3; Page 43-44; 63pp; English.
 PS The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX Sequence 286 AA;
 SQ

```
Query Match      68.2%; Score 73; DB 21; Length 286;
Best Local Similarity 68.4%; Pred. NO. 0.0019;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
   |||||:|:|:|:|:|
Db 60 NTTGVFGIEQWDRCVIS 78

RESULT 33
AAU04196
ID AAU04196 standard; Protein; 286 AA.
XX AC AAU04196;
XX DT 23-OCT-2001 (first entry)
XX DE Variable surface antigen 2 (VSA2) from Ehrlichia chaffeensis.
XX KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
XX KW infection; heartwater; diagnostic; variable surface antigen; VSA.
XX OS Ehrlichia chaffeensis.
XX PN US6251872-B1.
XX PD 26-JUN-2001.
XX PF 17-OCT-1997; 97US-0953326.
XX PR 17-OCT-1996; 96US-0733230.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX DR WPI; 2001-424487/45.
XX DR N-PSDB; AAS07578.
XX PT New MAP2 genes and polypeptides useful as vaccines for conferring
XX PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX PT as molecular markers in nucleic acid analysis procedures -
XX PS Example 3; Fig 2A-2B; 30pp; English.
XX CC The sequence represents the amino acid sequence of variable surface
XX CC antigen 2 (VSA2) isolated from Ehrlichia chaffeensis, which
XX CC has similarity to major antigen protein (MAP). The MAP polynucleotides
XX CC and polypeptides are useful as vaccines for conferring immunity to
XX CC rickettsia infection, including Cowdria ruminantium causing heartwater.
XX CC The MAP polynucleotides may be used as molecular markers in nucleic acid
XX CC analysis procedures, and to produce the MAP polypeptides, which may
XX CC be used to raise antibodies that are reactive with the polypeptides.
XX CC The nucleic acids may further be used as probes to identify
XX CC complementary sequences within other nucleic acid molecules or genomes,
XX CC where such probes can be applied to identify or distinguish infectious
XX CC strains of organisms in diagnostic procedures or in rickettsial
XX CC research where identification of particular organisms or strains is
XX CC needed.
XX SQ Sequence 286 AA;

Query Match      68.2%; Score 73; DB 22; Length 286;
Best Local Similarity 68.4%; Pred. NO. 0.0019;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
   |||||:|:|:|:|:|
Db 60 NTTGVFGIEQWDRCVIS 78

RESULT 34
```

```
ABG77938
ID ABG77938 standard; Protein; 286 AA.
XX AC ABG77938;
XX DT 15-NOV-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein (OMP) #4.
XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX OS Ehrlichia chaffeensis.
XX PN US2002120115-A1.
XX PD 29-AUG-2002.
XX PF 28-JAN-2002; 2002US-0059964.
XX PR 19-MAY-1999; 99US-0314701.
XX PA (RIKI/) RIKIHISA Y.
XX PA (OHAS/) OHASHI N.
XX PI Rikihisa Y, Ohashi N;
XX DR WPI; 2002-618954/66.
XX DR N-PSDB; ABS63279.
XX PT Isolated polynucleotide encoding an outer membrane protein of E.canis
XX PT or E.chaffeensis used in the diagnosis of infection -
XX PS Disclosure; Fig 6B; 49pp; English.
XX CC The invention relates to an isolated polynucleotide encoding an outer
XX CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX CC in the diagnosis of infection. An infection such as human ehrlichiosis or
XX CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX CC patient, providing a polypeptide or mixture of polypeptides, contacting
XX CC the sample with the polypeptide and assaying for the formation of a
XX CC complex between antibodies in the serum sample and the polypeptide, where
XX CC formation of a complex is indicative of infection with E. chaffeensis.
XX CC This sequence represents an Ehrlichia outer membrane protein of the
XX CC invention.
XX SQ Sequence 286 AA;

Query Match      68.2%; Score 73; DB 23; Length 286;
Best Local Similarity 68.4%; Pred. NO. 0.0019;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
   |||||:|:|:~|:|
Db 60 NTTGVFGIEQWDRCVIS 78

RESULT 35
AAU96108
ID AAU96108 standard; Protein; 286 AA.
XX AC AAU96108;
XX DT 02-JUL-2002 (first entry)
XX DE Ehrlichia chaffeensis OMP-1D.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX OS Ehrlichia chaffeensis.
XX PN WO200222782-A2.
```

```

PD 21-MAR-2002.
XX
XX
PF 12-SEP-2001; 2001WO-US28759.
XX
XX
PR 12-SEP-2000; 2000US-0660587.
XX
XX
PA (RERE-) RES DEV FOUND.
XX
XX
PI Walker DH, Yu X, McBride JW;
XX
XX
DR WPI; 2002-351882/38.
XX
XX
PT New recombinant homologous 28 kilodalton immunodominant protein from
XX Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX
XX
PS Example 3; Figure 3; 106pp; English.
XX
XX
CC The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
XX
SQ Sequence 286 AA;
Query Match 68.2%; Score 73; DB 23; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0019; 3; Indels 0; Gaps 0;
Matches 13; Conservative 3; Mismatches 3;
QY 1 NTTGVFGLKQNDGSAIS 19
Db |||||||:|:|:|
60 NTTGVFGIEODMDRCVIS 78

Search completed: October 6, 2003, 07:47:57
Job time : 36.3731 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 25.806 Seconds
(without alignments)
189.995 Million cell updates/sec

Title: US-09-765-739A-7

Perfect score: 103

Sequence: 1 NTTGVFLKQDWDGSTRIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_23:**

1: sp.archaea:**

2: sp.bacteria:**

3: sp.fungi:**

4: sp.human:**

5: sp.invertebrate:**

6: sp.mammal:**

7: sp.mhc:**

8: sp.organelle:**

9: sp.phage:**

10: sp.plant:**

11: sp.rodent:**

12: sp.virus:**

13: sp.vertibrate:**

14: sp.unclassified:**

15: sp.virus:**

16: sp.bacteriap:**

17: sp.archeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	103	100.0	280	2	052107	052107 ehrlichia c
2	96	93.2	288	2	Q9ZGJ2	Q9ZGJ2 ehrlichia c
3	89	86.4	246	2	Q9RH35	Q9RH35 ehrlichia c
4	89	86.4	275	2	Q93DD4	Q93DD4 ehrlichia c
5	89	86.4	276	2	Q93DD1	Q93DD1 ehrlichia c
6	89	86.4	276	2	Q95817	Q95817 ehrlichia c
7	89	86.4	276	2	Q8GGU0	Q8GGU0 ehrlichia c
8	88	85.4	280	2	Q93DD3	Q93DD3 ehrlichia c
9	88	85.4	280	2	Q9ZGM9	Q9ZGM9 ehrlichia c
10	88	85.4	280	2	Q95816	Q95816 ehrlichia c
11	88	85.4	281	2	Q93DD2	Q93DD2 ehrlichia c
12	88	85.4	281	2	Q9AC19	Q9AC19 ehrlichia c
13	80	77.7	280	2	Q9ADV3	Q9ADV3 ehrlichia c
14	80	77.7	280	2	Q9F473	Q9F473 ehrlichia c
15	74	71.8	286	2	052105	052105 ehrlichia c
16	74	71.8	287	2	Q8GGU3	Q8GGU3 ehrlichia c

Q8GGU2 ehrlichia c
Q8G921 ehrlichia c
Q9F472 ehrlichia c
Q9R8A8 ehrlichia c
Q9R8A7 ehrlichia c
Q9R3J3 ehrlichia c
Q9R8A6 ehrlichia c
Q9R8A9 ehrlichia c
Q9R8A5 ehrlichia c
Q9ZGJ1 ehrlichia c
Q9ZGJ1 ehrlichia c
Q9F475 ehrlichia c
Q8G8P3 ehrlichia c
Q8G948 ehrlichia c
Q8GGU1 ehrlichia c
Q8G8W7 ehrlichia c
Q52106 ehrlichia c
Q52104 ehrlichia c
Q8GGU5 ehrlichia c
Q8G8J3 ehrlichia c
Q9AF99 cowdria rum
Q9AF98 cowdria rum
Q93E59 cowdria rum
Q93E60 cowdria rum
Q93E53 cowdria rum
Q93E52 cowdria rum
Q93E57 cowdria rum
Q93E56 cowdria rum
Q46329 cowdria rum
Q9R425 cowdria rum
Q46331 cowdria rum
Q46324 cowdria rum
Q93E64 cowdria rum
Q46333 cowdria rum
Q46332 cowdria rum
Q46330 cowdria rum
Q9AMF6 ehrlichia s
Q93E54 cowdria rum
Q9AF41 cowdria rum
Q46327 cowdria rum
Q9AEU3 cowdria rum
Q93E62 cowdria rum
Q9S6H1 cowdria rum
Q46328 cowdria rum
Q9S6H0 cowdria rum
Q8S360 ehrlichia c
Q93E55 cowdria rum
Q93E58 cowdria rum
Q9F474 ehrlichia c
Q9AF40 cowdria rum
Q9FKI3 arabisopsis
Q818M9 arabisopsis
Q93E61 cowdria rum
Q93E63 cowdria rum
Q8X4H5 escherichia
Q8X2T1 escherichia
Q06352 escherichia
Q8XQP7 ralsionia s
Q929J3 listeria in
Q9U993 acanthochei
Q9Y0A5 acanthochei
Q9W3R8 drosophila
Q8P3Y5 xanthomonas
Q90423 vibrio chol
Q9K422 brachydanio
Q9CFA2 lactococcus
Q9HZ01 pseudomonas
Q8TI54 methanosarc
Q9SWL1 lycopersico
Q02659 saccharomyc
Q8QAW8 human immun
Q9M0C6 arabisopsis

90 43 41.7 498 10 Q9FHH4 Q9fhh4 arabidopsis
91 43 41.7 623 10 Q9LU26 Q9lu26 arabidopsis
92 43 41.7 806 16 Q8NNL3 Q8nnl3 corynabacte
93 43 41.7 855 16 Q8FNR9 Q8fnr9 corynabacte
94 43 41.7 904 10 Q9FU09 Q9fu09 phaseolus v
95 43 41.7 926 10 Q9ZSH1 Q9zsh1 brassica na
96 43 41.7 1429 5 Q9W5A4 Q9w5a4 drosophila
97 42 40.8 79 9 Q9B041 Q9b041 mycobacteri
98 42 40.8 160 16 Q8YLD3 Q8yld3 anabaena sp
99 42 40.8 307 16 Q8VLS0 Q8vls0 anabaena sp
100 42 40.8 347 15 Q8QAW5 Q8qaw5 human immun

ALIGNMENTS

RESULT 1
ID Q52107 PRELIMINARY; PRT; 280 AA.
AC Q52107; (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Major outer membrane protein OMP-1F (28kDa outer membrane protein gene
DE 18).
GN OMP-1F.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-133(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
RN [5]
DR EMBL; U72291; AAC02940.1; -;
DR EMBL; AF479833; AA012931.1; -;
DR EMBL; AF479834; AA012937.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAAGC34E2AF393E CRC64;

Query Match Score 100.0%; DB 2; Length 280;

Best Local Similarity 100.0%; Pred. No. 3.9e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NNTTGVFLKQDWGSGTIS 19
DB 60 NNTTGVFLKQDWGSGTIS 78
RESULT 2
Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
ID Q9ZGJ2
AC Q9ZGJ2;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 30-kDa major outer membrane protein (P28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jaks;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Jaks;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL; AF078553; AAC68667.1; -;
DR EMBL; AF082744; AAG14362.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BF5E CRC64;
Query Match Score 93.2%; DB 2; Length 288;
Best Local Similarity 94.4%; Pred. No. 5.1e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 NNTTGVFLKQDWGSGTI 18
DB 60 NNTTGVFLKQDWGSGTI 77
RESULT 3
Q9RH35 PRELIMINARY; PRT; 246 AA.
ID Q9RH35
AC Q9RH35;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Outer membrane protein p28 (Fragment).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

```
OC Anaplasmataceae; Ehrlichia.
RN NCBI_TaxID=945;
RP SEQUENCE FROM N.A.
RC STRAIN=St.Vincent;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 86.4%; Score 89; DB 2; Length 246;
Best Local Similarity 88.9%; Pred. No. 5.5e-06;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSGTI 18
DB 29 NTTAGVFGKQDWGSAI 46

RESULT 4
Q3DD4 PRELIMINARY; PRT; 275 AA.
AC Q3DD4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V2;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393389; AAL12919.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 29974 MW; 2ECCF2F988B2E9D9 CRC64;

Query Match 86.4%; Score 89; DB 2; Length 275;
Best Local Similarity 88.9%; Pred. No. 6.2e-06;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSGTI 18
DB 59 NTTAGVFGKQDWGSAI 76

RESULT 5
Q3DD1 PRELIMINARY; PRT; 276 AA.
AC Q3DD1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393393; AAL12925.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30028 MW; 2D7143AFCBF2EBE CRC64;

Query Match 86.4%; Score 89; DB 2; Length 276;
Best Local Similarity 88.9%; Pred. No. 6.2e-06;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGSGTI 18
DB 59 NTTAGVFGKQDWGSAI 76

RESULT 6
Q85817 PRELIMINARY; PRT; 276 AA.
AC Q85817;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28 kDa outer membrane protein).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V9;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Chattanooga, West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF077734; AAC31547.1; -.
DR EMBL: AF393395; AAL12925.1; -.
DR EMBL: AV117397; AAM77032.1; -.
DR EMBL: AF479835; AAO12943.1; -.
DR EMBL: AF479836; AAO12948.1; -.
DR EMBL: AF479837; AAO12953.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 86.4%; Score 89; DB 2; Length 276;
Best Local Similarity 88.9%; Pred. No. 6.2e-06;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 NTTGTGVLKQDWDGSKI 18
DB 59 NTTAGVFLKQDWDGSAI 76

RESULT 7
Q8GGU0 PRELIMINARY; PRT; 276 AA.
AC Q8GGU0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 28KDa outer membrane protein gene 19.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wakulla;
RX PubMed=12496165;
RA "Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479838; AA012958.1; -.
SQ SEQUENCE 276 AA; 30045 MW; 8E3C9719D3C67A64 CRC64;

Query Match 86.4%; Score 89; DB 2; Length 276;
Best Local Similarity 88.9%; Pred. No. 6.2e-06;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGTGVLKQDWDGSKI 18
DB 59 NTTAGVFLKQDWDGSAI 76

RESULT 8
Q93DD3 PRELIMINARY; PRT; 280 AA.
AC Q93DD3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V5;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393391; AAL12921.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30372 MW; C7B8C8710BC167E9 CRC64;

Query Match 85.4%; Score 88; DB 2; Length 280;
Best Local Similarity 84.2%; Pred. No. 9.1e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGTGVLKQDWDGSKI 19
DB 59 STTAGVFLKQDWDGSAIS 77

RESULT 9
Q9ZGM9 PRELIMINARY; PRT; 280 AA.
AC Q9ZGM9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Query Match 85.4%; Score 88; DB 2; Length 280;
Best Local Similarity 84.2%; Pred. No. 9.1e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGTGVLKQDWDGSKI 19
DB 59 STTAGVFLKQDWDGSAIS 77

RESULT 10
Q85816 PRELIMINARY; PRT; 280 AA.
AC Q85816;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28kDa outer membrane protein gene
DE 19).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V8, and V4;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
```

RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077732; AAC31545.1; -
DR EMBL; AF393394; AAL12924.1; -
DR EMBL; AF393390; AAL12920.1; -
DR EMBL; AF479839; AAO12964.1; -
DR EMBL; AF479840; AAO12970.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 85.4%; Score 88; DB 2; Length 280;
Best Local Similarity 84.2%; Pred. No. 9.1e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGSGTIS 19
:|| ||||| ||||| ||
Db 59 STTAGVFLKQDWDGSAIS 77

RESULT 11

Q93DD2 PRELIMINARY; PRT; 281 AA.
AC Q93DD2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V6;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393392; AAL12922.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30388 MW; B54DBB55AC839A9A CRC64;

Query Match 85.4%; Score 88; DB 2; Length 281;
Best Local Similarity 84.2%; Pred. No. 9.1e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGSGTIS 19
||| ||||| ||||| ||
Db 59 NTTTGVFLKQDWDGSAIS 77

RESULT 12

Q9AC19 PRELIMINARY; PRT; 281 AA.
AC Q9AC19;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Major outer membrane protein p28 (28 kDa outer membrane protein).
GN P28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;

RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V1;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Lithonia, Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=96321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; U72291; AAK28673.1; -
DR EMBL; AF393388; AAL12918.1; -
DR EMBL; AF117396; AAM77031.1; -
DR EMBL; AF479833; AAO12932.1; -
DR EMBL; AF479834; AAO12938.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 85.4%; Score 88; DB 2; Length 281;
Best Local Similarity 84.2%; Pred. No. 9.1e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWDGSGTIS 19
||| ||||| ||||| ||
Db 59 NTTTGVFLKQDWDGSAIS 77

RESULT 13

Q9ADV3 PRELIMINARY; PRT; 280 AA.
AC Q9ADV3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE Major outer membrane protein p30-2.
GN P30-2.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;

```

RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RA MEDLINE=21153566; PubMed=11254561;
RX Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis."; 69:2083-2091(2001).
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAK28699.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match 77.7%; Score 80; DB 2; Length 280;
Best Local Similarity 73.7%; Pred. No. 0.00016;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWDGSGTIS 19
Db i:|||||i|||i|||

59 NSTGVFGLKHDWNGGTIS 77

RESULT 14
ID Q9F473 PRELIMINARY; PRT; 280 AA.
AC Q9F473;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-6.
GN Ehrlichia canis.
OS Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RA MEDLINE=99242757; PubMed=10225842;
RX McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RA MEDLINE=20432107; PubMed=10974556;
RX McBride J.W., Yu, X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis."; 2000.
RL Gene 234:245-252(2000).
DR EMBL; AF082744; AAC14361.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match 77.7%; Score 80; DB 2; Length 280;
Best Local Similarity 73.7%; Pred. No. 0.00016;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWDGSGTIS 19
Db i:|||||i|||i|||

59 NSTGVFGLKHDWNGGTIS 77

```

```

RESULT 15
ID O52105 PRELIMINARY; PRT; 286 AA.
AC O52105;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OMP-1D (28kDa outer membrane protein gene 16).
GN OMP-1D.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA MEDLINE=98084465; PubMed=9423849;
RX Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.;
RX Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Oscicola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02938.1; -.
DR EMBL; AF479833; AAC26718.1; -.
DR EMBL; AF479834; AAC12935.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 286 AA; 31509 MW; F145A79270F386BE CRC64;

Query Match 71.8%; Score 74; DB 2; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWDGSGTIS 19
Db ||| ||||:|||||

60 NTTGVFGIEQDMDRCVIS 78

RESULT 16
ID Q8GGU3 PRELIMINARY; PRT; 287 AA.
AC Q8GGU3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;

```

RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479837; AAO12951.1; -.
SQ SEQUENCE 287 AA; 31504 MW; A9592249C83695B2 CRC64;

Query Match 71.8%; Score 74; DB 2; Length 287;
Best Local Similarity 68.4%; Pred. No. 0.0015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
||| |||||:|||| ||
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 17
Q8GGU2 PRELIMINARY; PRT; 291 AA.
AC Q8GGU2;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479838; AAO12956.1; -.
SQ SEQUENCE 291 AA; 32005 MW; 22D35EDED283195A CRC64;

Query Match 71.8%; Score 74; DB 2; Length 291;
Best Local Similarity 68.4%; Pred. No. 0.0015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
||| |||||:|||| ||
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 18
Q8G921 PRELIMINARY; PRT; 291 AA.
AC Q8G921;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-West Paces, and Heartland;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479835; AAO12941.1; -.
DR EMBL: AF479836; AAO12946.1; -.
SQ SEQUENCE 291 AA; 31991 MW; 22D35A2202831369 CRC64;

Query Match 71.8%; Score 74; DB 2; Length 291;
Best Local Similarity 68.4%; Pred. No. 0.0015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
||| |||||:|||| ||
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 19
Q9F472 PRELIMINARY; PRT; 278 AA.
AC Q9F472;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE P28-7.
GN P28-7.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL: AF082744; AAC64550.2; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2.1.
SQ SEQUENCE 278 AA; 30455 MW; 2411CAAB4C56CA74 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWGSGTI 18
:| |||||:|||| |
Db 61 STGVFGLKQDWDGSP 77

RESULT 20
Q9R8A8 PRELIMINARY; PRT; 278 AA.
ID Q9R8A8
AC Q9R8A8;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Okiahoma;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic

RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082746; AAC64552.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag-2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWDGSTI 18
: | | | | | | | | | |
Db 61 STVGVEGLKHDWDGSP 77

RESULT 21

Q9R8A7 PRELIMINARY; PRT; 278 AA.
AC Q9R8A7;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Demon;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082747; AAC64553.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag-2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWDGSTI 18
: | | | | | | | | | |
Db 61 STVGVEGLKHDWDGSP 77

RESULT 22

Q9R3J3 PRELIMINARY; PRT; 278 AA.
AC Q9R3J3;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Florida;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";

RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082750; AAC64556.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag-2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWDGSTI 18
: | | | | | | | | | |
Db 61 STVGVEGLKHDWDGSP 77

RESULT 23

Q9R8A6 PRELIMINARY; PRT; 278 AA.
AC Q9R8A6;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DJ;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082748; AAC64554.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag-2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWDGSTI 18
: | | | | | | | | | |
Db 61 STVGVEGLKHDWDGSP 77

RESULT 24

Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Louisiana;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).

DR EMBL; AF082745; AAC64551.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWGSTI 18
:|||||||
Db 61 STVGVFGLKHDWDGSP I 77

RESULT 25

Q9R8A5 PRELIMINARY; PRT; 278 AA.
ID Q9R8A5;
AC Q9R8A5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE 28 kda outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Fuzzy;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, XJ, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082749; AAC64555.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWGSTI 18
:|||||||
Db 61 STVGVFGLKHDWDGSP I 77

RESULT 26

Q9ZGJ1 PRELIMINARY; PRT; 307 AA.
ID Q9ZGJ1;
AC Q9ZGJ1;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Major outer membrane protein P30-1.
GN P30-1.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).

RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF078553; AAC68666.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 307 AA; 34108 MW; 9DA9FD63EBF8BC97 CRC64;

Query Match 70.9%; Score 73; DB 2; Length 307;
Best Local Similarity 76.5%; Pred. No. 0.0023;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFGLKQDWGSTI 18
:|||||||
Db 90 STVGVFGLKHDWDGSP I 106

RESULT 27

Q9F475 PRELIMINARY; PRT; 276 AA.
ID Q9F475
AC Q9F475;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE P28-3 (Major outer membrane protein P30-4).
GN P28-3 OR P30-4.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, XJ, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).

QY 2 TTTGVFGLKQDWGSTI 18
:|||||||
Db 90 STVGVFGLKHDWDGSP I 106

Q9F475 PRELIMINARY; PRT; 276 AA.
ID Q9F475
AC Q9F475;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE P28-3 (Major outer membrane protein P30-4).
GN P28-3 OR P30-4.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, XJ, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).

DR EMBL: AF082744; AAC14358.1; -;
DR EMBL: AF324792; AAK31313.1; -;
DR EMBL: AF078553; AAK28697.1; -;
DR InterPro: IPR002366; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30659 MW; CE51AB37D17AF3A4 CRC64;

Query Match 69.9%; Score 72; DB 2; Length 276;
Best Local Similarity 66.7%; Pred. No. 0.003;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWDGSTI 18
| | | | | | | | | | | | | | | | | |
DB 60 NTTTGIFGLKESWTGGII 77

RESULT 28

O8G8P3 PRELIMINARY; PRT; 291 AA.
AC O8G8P3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of the Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479839; AAO12962.1; -;
DR EMBL: AF479840; AAO12968.1; -;
SQ SEQUENCE 291 AA; 32130 MW; 8220C436A1FC08E2 CRC64;

Query Match 67.0%; Score 69; DB 2; Length 291;
Best Local Similarity 63.2%; Pred. No. 0.0093;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWDGSTIS 19
| | | | | | | | | | | | | | | | | |
DB 60 NTTTGVFGLKQDWNRCVIS 78

RESULT 29

O8G948 PRELIMINARY; PRT; 276 AA.
AC O8G948;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE 28kDa outer membrane protein gene 15.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=West Paces, Heartland, St. Vincent, and Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479835; AAO12940.1; -;
DR EMBL: AF479836; AAO12945.1; -;

DR EMBL: AF479837; AAO12950.1; -;
DR EMBL: AF479838; AAO12955.1; -;
SQ SEQUENCE 276 AA; 30316 MW; OD6F5353F9C0F17C CRC64;

Query Match 62.1%; Score 64; DB 2; Length 276;
Best Local Similarity 52.6%; Pred. No. 0.054;
Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWDGSTIS 19
| | | | | | | | | | | | | | | | | |
DB 60 NATVALYGLKQDNGASAS 78

RESULT 30

O8GGU1 PRELIMINARY; PRT; 277 AA.
AC O8GGU1;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479838; AAO12957.1; -;
SQ SEQUENCE 277 AA; 30294 MW; 0CE7EDB51F2D854E CRC64;

Query Match 62.1%; Score 64; DB 2; Length 277;
Best Local Similarity 52.6%; Pred. No. 0.054;
Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWDGSTIS 19
| | | | | | | | | | | | | | | | | |
DB 60 NATVAMYGLKQDNGASAS 78

RESULT 31

O8G8W7 PRELIMINARY; PRT; 277 AA.
AC O8G8W7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479835; AAO12942.1; -;
DR EMBL: AF479836; AAO12947.1; -;
DR EMBL: AF479837; AAO12952.1; -;
SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

Query Match 62.1%; Score 64; DB 2; Length 277;

Best Local Similarity 52.6%; Pred. No. 0.054;
Matches 10; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
| | :|||||:|:| |
Db 60 NATVALYGLKQDWNGASAS 78

RESULT 32

OS2106 PRELIMINARY; PRT; 278 AA.
AC OS2106;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE OMP-1E (28kDa outer membrane protein gene 17).
GN OMP-1E
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02939.1; -;
DR EMBL; AF479833; AAC26719.1; -;
DR EMBL; AF479834; AAO12936.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2_1.
SQ SEQUENCE 278 AA; 30543 MW; E321E3CA259B87FD CRC64;

Query Match 58.3%; Score 60; DB 2; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.23;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
| | :|||||:|:| |
Db 60 NPTVALYGLKQDWEGISSS 78

RESULT 33

Q8G8Q5 PRELIMINARY; PRT; 278 AA.
AC Q8G8Q5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Liberty; and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479839; AAO12963.1; -;
DR EMBL; AF479840; AAO12969.1; -;
SQ SEQUENCE 278 AA; 30381 MW; 7626778E5798D6B7 CRC64;

Query Match 58.3%; Score 60; DB 2; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.23;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGSGTIS 19
| | :|||||:|:| |
Db 60 NPTVALYGLKQDWEGISSS 78

RESULT 34

OS2104 PRELIMINARY; PRT; 280 AA.
AC OS2104; O85357;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE OMP-1C (28kDa outer membrane protein 15).
GN OMP-1C
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ARKANSAS;
RX MEDLINE=22384137; PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02937.1; -;
DR EMBL; AF479833; AAO12930.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2_1.
SQ SEQUENCE 280 AA; 30322 MW; BD835E491086DF01 CRC64;

Query Match 58.3%; Score 60; DB 2; Length 280;
Best Local Similarity 52.6%; Pred. No. 0.23;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSGTIS 19
 Db 60 NPTVALYGLKQDWNGVSAS 78

RESULT 35

Q8GGU5 PRELIMINARY; PRT; 280 AA.
 AC Q8GGU5;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 28kDa outer membrane protein gene 15.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oscicola;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 RT by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
 RT and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479834; AAO12934.1; -.
 SQ SEQUENCE 280 AA; 30332 MW; BD835D792386DF01 CRC64;

Query Match 58.3%; Score 60; DB 2; Length 280;
 Best Local Similarity 52.6%; Pred. NO. 0.23;
 Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSGTIS 19
 Db 60 NPTVALYGLKQDWNGVSAS 78

Search completed: October 6, 2003, 07:52:43
 Job time : 27.806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:02 ; Search time 5.38806 Seconds
(without alignments)
165.831 Million cell updates/sec

Title: US-09-765-739A-7

Perfect score: 103

Sequence: 1 NTTTTGFLKQDWDGSTITIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	51	49.5	540	1 YH8M_YEAST	Q03263 saccharomyc
2	43.5	42.2	290	1 AROE_SYNT3	P74591 synechocyst
3	43	41.7	160	1 MENG_DEIRA	Q9rlw0 deinococcus
4	43	41.7	648	1 VP3_BPPH6	P11129 bacterioph
5	42.5	41.3	734	1 PURL_ZYMO	Q9req6 zymomonas m
6	41	39.8	135	1 FARE_BOVIN	P55052 bos taurus
7	41	39.8	135	1 FARE_HUMAN	Q01469 homo sapien
8	41	39.8	689	1 SYGB_PASMO	P57905 pasteurella
9	41	39.8	866	1 YCBS_ECOLI	P75857 escherichia
10	40	38.8	115	1 GCSH_CHLPN	Q928b0 chlamydia p
11	40	38.8	180	1 RBS_MARPA	O64416 marchantia
12	40	38.8	282	1 BIOB_HELPJ	Q9zrk8 helicobacte
13	40	38.8	282	1 BIOB_HELPJ	O25956 helicobacte
14	40	38.8	377	1 YA67_METH	O27139 methanobact
15	40	38.8	473	1 SYE1_BRUME	Q8yhq4 brucella su
16	40	38.8	473	1 SYE1_BRUSE	Q8yhq4 brucella su
17	40	38.8	475	1 MTHC_DROME	P83119 drosophila
18	40	38.8	741	1 PLO3_MOUSE	Q9r0e1 mus musculu
19	40	38.8	942	1 ENV_CAEVG	P31627 caprine art
20	40	38.8	1025	1 DPYD_HUMAN	Q12882 homo sapien
21	40	38.8	1191	1 S122_SQAC	P55013 squallus aca
22	39.5	38.3	366	1 GCST_NEIMA	Q9k018 neisseria m
23	39.5	38.3	368	1 GCST_NEIMA	Q9jvp2 neisseria m
24	39	37.9	149	1 DTD_CLOAB	Q97gu2 clostridium
25	39	37.9	188	1 EFPL_XYLET	Q87c43 xylella fas
26	39	37.9	189	1 EFPL_XYLET	Q87c43 xylella fas
27	39	37.9	282	1 PUNA_CELSP	P81989 cellulomona
28	39	37.9	284	1 YQAK_BACSU	P45908 bacillus su
29	39	37.9	285	1 SYGB_HAREE	Q30836 haemophilus
30	39	37.9	298	1 HEM3_FUSNN	O8rfp5 fusobacteri
31	39	37.9	307	1 YML6_MYCLE	O32960 mycobacteri
32	39	37.9	326	1 GBLP_NICPL	P93340 nicotiana p
33	39	37.9	326	1 GBLP_TOBAC	P49026 nicotiana t

ALIGNMENTS

RESULT 1

34	37.9	422	1 YF02_MYCPN	P75285 mycoplasma
35	37.9	427	1 PYRC_BACCL	P45338 bacillus ca
36	37.9	428	1 YB01_MYCPN	P75568 mycoplasma
37	37.9	438	1 YC05_MYCPN	P75571 mycoplasma
38	37.9	471	1 SYE_RHOSH	Q9zra3 rhodobacter
39	37.9	473	1 XYL_A_CLOS	P48790 clostridium
40	37.9	688	1 SYGB_HAEIN	P43822 haemophilus
41	37.9	756	1 KGPF_CAEEL	Q27483 caenorhabdi
42	37.9	896	1 POL_HPLIC	P14078 human t-cel
43	37.9	1004	1 ARIB_ARTSF	P28774 artemia san
44	37.9	1025	1 DPYD_BOVIN	Q28007 bos taurus
45	37.9	1025	1 DPYD_PIG	Q28943 sus scrofa
46	37.9	1203	1 YF41_CAEEL	Q11069 caenorhabdi
47	37.9	1586	1 ARO1_EMENI	P07547 e pentafunc
48	37.9	536	1 FYN_XIPHE	P27446 xiphophorus
49	37.9	197	1 VG23_BPMD2	P38362 mycobacteri
50	37.9	233	1 RR2_CYAPA	P48132 cyanophora
51	37.9	298	1 YMA2_MYCBO	Q02278 mycobacteri
52	37.9	322	1 PTNA_ECOLI	P08186 escherichia
53	37.9	341	1 Y33B_MYCPN	P75302 mycoplasma
54	37.9	371	1 Y028_ARCFU	O30207 archaeoglob
55	37.9	372	1 RENG_CHICK	O12972 gallus gall
56	37.9	406	1 YNO5_YEAST	P53891 saccharomyc
57	37.9	472	1 PEDA_STREY	Q940m0 streptococc
58	37.9	508	1 PSBB_CHLRE	P37255 chlamydomon
59	37.9	508	1 PSBB_CHLVU	P56307 chlorella v
60	37.9	510	1 GUAA_CLOAB	O97fm9 clostridium
61	37.9	558	1 CSG_METJA	Q58232 methanococc
62	37.9	610	1 CC20_YEAST	P26309 saccharomyc
63	37.9	669	1 VPH_BPP2	P25700 bacterioph
64	37.9	697	1 SIIL_HUMAN	Q961c7 homo sapien
65	37.9	749	1 VP4_ROTGA	Q04916 rotavirus (
66	37.9	959	1 N100_YEAST	Q02629 saccharomyc
67	37.9	1021	1 AIAL_CANE	P50997 canis fami
68	37.9	1182	1 CGA2_HELPY	P53746 helicobacte
69	37.9	2376	1 TAO3_YEAST	P40468 saccharomyc
70	37.5	419	1 CHSD_PETHY	P22925 petunia hyb
71	37.5	684	1 HTPG_PORGI	Q983q2 porphyromon
72	37.5	84	1 GVM1_HALNI	P24377 halobacteri
73	37.5	191	1 PGHD_URSAR	Q29562 ursus arcto
74	37.5	196	1 NODA_MESPL	Q8vvh1 mesorhizobi
75	37.5	289	1 PORI_RHOBL	P97671 rhodopseudo
76	37.5	317	1 GBLP_BRARE	O42248 brachydanio
77	37.5	317	1 GBLP_HUMAN	P25388 homo sapien
78	37.5	341	1 EFG_STPRA	P29541 streptomyc
79	37.5	347	1 NU2M_BALMU	P41297 balaeopter
80	37.5	353	1 VM17_BORHE	P32777 borrelia he
81	37.5	362	1 QUEA_DEIRA	Q9rlu9 deinococcus
82	37.5	370	1 ARIA_HUMAN	Q92747 homo sapien
83	37.5	395	1 KIME_RAT	P17256 rattus norv
84	37.5	407	1 GSPL_VIBCH	P45782 vibrio chol
85	37.5	429	1 YVFO_BACSU	O07013 bacillus su
86	37.5	463	1 FLGE_TREPH	Q56326 treponema p
87	37.5	465	1 YHJA_ECOLI	P37197 escherichia
88	37.5	500	1 AMPA_BACSU	O32106 bacillus su
89	37.5	523	1 GUAA_CORGL	Q8nsr1 corynebacte
90	37.5	704	1 DPOL_BPT3	P20311 bacterioph
91	37.5	704	1 DPOL_BPT7	P00581 bacterioph
92	37.5	708	1 EFG1_STRCO	P40173 streptomyc
93	37.5	738	1 SMO3_HUMAN	O60568 homo sapien
94	37.5	770	1 P404_HUMAN	O59754 homo sapien
95	37.5	802	1 XYND_RUMFL	Q53317 ruminococcu
96	37.5	863	1 YEJO_ECOLI	P33924 escherichia
97	37.5	1020	1 AIA2_HUMAN	P50993 homo sapien
98	37.5	1034	1 CAPP_SYNT3	P74299 synechocyst
99	37.5	1699	1 POLN_LORDV	P54634 lordsdale v
100	37.5	2126	1 PKDR_MOUSE	Q920t6 mus musculu

```
YMBM_YEAST          STANDARD;          PRT;          540 AA.
ID  Q03263;
AC  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  Hypothetical 59.6 kDa protein in DSK2-CAT8 intergenic region.
GN  YMR279C OR YMR021.05C.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=97061201; PubMed=8905231;
RX  Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA  Miyaajima N., Hirotsawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA  Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA  Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA  Tabata S.,
RT  *Sequence analysis of the genome of the unicellular cyanobacterium
RT  Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT  entire genome and assignment of potential protein-coding regions.*;
RL  DNA Res. 3:109-136(1996).
CC  -!- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate +
CC  NADPH.
CC  -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC  fourth step.
CC  -!- SIMILARITY: Belongs to the shikimate dehydrogenase family.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D90916; BAA18699.1; -
DR  PIR; S76787; S76787.
DR  HAMAP; MF_00222; -; 1.
DR  InterPro; IPR006151; Shikimate_DH.
DR  Pfam; PF01488; Shikimate_DH; 1.
DR  TIGRFAMs; TIGR00507; aroe; 1.
KW  Aromatic amino acid biosynthesis; Oxidoreductase; NADP;
KW  Complete proteome.
SQ  SEQUENCE 290 AA; 31099 MW; 8A2D38EE5D57B303 CRC64;
Query Match 42.2%; Score 43.5; DB 1; Length 290;
Best Local Similarity 47.8%; Pred. No. 8.7;
Matches 11; Conservative 0; Mismatches 5; Indels 7; Gaps 1;
OY 2 TTTGVFG-----LKQDWDGST 17
DB 105 TNDTVEGFLAPLLELQKQDSGRT 127
RESULT 3
MENG_DEIRA
ID MENG_DEIRA STANDARD; PRT; 160 AA.
AC Q9RW10;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable S-adenosylmethionine:2-demethylmenaquinone methyltransferase
DE (EC 2.1.1.-).
GN MENG OR DR0859.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567366;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
Shikimate 5-dehydrogenase (EC 1.1.1.25).
DE AROE OR SLR1559.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97061201; PubMed=8905231;
RX Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyaajima N., Hirotsawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.,
RT *Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.*;
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate +
CC NADPH.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC fourth step.
CC -!- SIMILARITY: Belongs to the shikimate dehydrogenase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC 249704; CA89777.1; -
DR PIR; S54386; S54386.
DR SGD; S0004892; YMR279C.
DR InterPro; IPR007114; MFS.
KW Hypothetical protein; Transport; Transmembrane.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 109 129 POTENTIAL.
FT TRANSMEM 132 152 POTENTIAL.
FT TRANSMEM 170 190 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 430 450 POTENTIAL.
FT TRANSMEM 462 482 POTENTIAL.
FT TRANSMEM 503 523 POTENTIAL.
SQ SEQUENCE 540 AA; 59561 MW; 687D06CB0D70AF91 CRC64;
Query Match 49.5%; Score 51; DB 1; Length 540;
Best Local Similarity 42.1%; Pred. No. 0.98;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
OY 1 NNTTGVFGLKQDWDGSTRIS 19
DB 256 NVPTNIHGLSMDWTGSALA 274
RESULT 2
AROE_SYNY3
ID AROE_SYNY3 STANDARD; PRT; 290 AA.
AC P74591;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```

```
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: CONVERTS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK)
CC (By similarity).
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE MENG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001940; AAF10437.1; -.
DR PIR; A75466; A75466.
DR TIGR; DR0859; -.
DR HAMAP; MF_00471; -.
DR InterPro; IPR005493; Methyltransf_6.
DR Pfam; PF03737; Methyltransf_6; 1.
KW Menaquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 160 AA; 16860 MW; 54D6F226CA19EC0E CRC64;

Query Match 41.7%; Score 43; DB 1; Length 160;
Best Local Similarity 50.0%; Pred. No. 5.5;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGKQDWGSTI 18
DQ 80 GVFGVNGEGVII 93
DB

RESULT 4
VP3_BPPH6 STANDARD; PRT; 648 AA.
AC P11129;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DE P3 protein.
GN P3
OS Bacteriophage phi-6.
OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
OX NCBI_TaxID=10879;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.
RX MEDLINE=88160044; PubMed=3347997;
RA Gottlieb P., Metzger S., Romantschuk M., Carton J., Strassman J.,
RA Bamford D.H., Kaikkinen N., Mindich L.;
RT "Nucleotide sequence of the middle dsRNA segment of bacteriophage phi
RT 6: placement of the genes of membrane-associated proteins.";
RL Virology 163:183-190(1988).
CC -!- FUNCTION: P3 PROTEIN IS NECESSARY FOR ADSORPTION ONTO HOST
CC CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17462; AAA68485.1; -.
DR PIR; C28648; P3BPF6.
KW Envelope protein.
SQ SEQUENCE 648 AA; 69178 MW; B188DFE02ACC54E3 CRC64;
```

```
Query Match 41.7%; Score 43; DB 1; Length 648;
Best Local Similarity 43.8%; Pred. No. 25;
Matches 7; Conservative 5; Mismatches 2; Indels 2; Gaps 1;

QY 6 VFG--LKQDWDGSTIS 19
DQ 181 IFGVYVKMDWEGSAVA 196
DB

RESULT 5
PURL_ZYMMO STANDARD; PRT; 734 AA.
ID PURL_ZYMMO
AC Q9REQ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DE synthase II).
GN PUR-Q.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 31821 / ZM4 / CP4;
RA Um H.W., Kang H.S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
CC ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
CC (formamido)-N(1)-(5-phospho-D-ribosyl)acetamidine + L-glutamate.
CC -!- PATHWAY: De novo purine biosynthesis; fourth step.
CC -!- SUBUNIT: Heterodimer of two subunits, purQ and purL.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF213822; AAF23789.1; -.
DR HAMAP; MF_00420; 1.
DR InterPro; IPR000728; AIRS-related.
DR Pfam; PF00586; AIRS; 2.
DR Pfam; PF02769; AIRS_C; 2.
KW Purine biosynthesis; Ligase; ATP-binding.
FT NP_BIND 106 117
SQ SEQUENCE 734 AA; 77679 MW; B771635E0F66A166 CRC64;

Query Match 41.3%; Score 42.5; DB 1; Length 734;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 TTGVFLKQDW-DGSTIS 19
DQ 551 TIGVGVLQDWRDSTTIA 568
DB

RESULT 6
FABE_BOVIN STANDARD; PRT; 135 AA.
ID FABE_BOVIN
AC P55052; O62808;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Fatty acid-binding protein, epidermal (E-FABP) (Differentiation-
DE associated lipid binding protein LP2).
GN FABP5.
```

OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovinae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-33 AND 116-129.
 RC TISSUE=Leads;
 RX MEDLINE=97103094; PubMed=8947466;
 RA Jaworski C., Wistow G.;
 RT "Lp2, a differentiation-associated lipid-binding protein expressed in
 RT bovine lens.";
 RL Biochem. J. 320:49-54(1996).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Retina;
 RX MEDLINE=98198033; PubMed=9521644;
 RA Kingma P.B., Bok D., Ong D.E.;
 RT "Bovine epidermal fatty acid-binding protein: determination of ligand
 RT specificity and cellular localization in retina and testis.";
 RL Biochemistry 37:3250-3257(1998).
 CC -1- FUNCTION: HIGH SPECIFICITY FOR FATTY ACIDS. HIGHEST AFFINITY FOR
 CC THE UNSATURATED FATTY ACID PALMITATE (C16:0). MAY BE INVOLVED IN
 CC THE UPTAKE AND TRANSPORT OF FATTY ACIDS ESSENTIAL FOR THE
 CC NOURISHMENT OF THE SURROUNDING CELL TYPES. IT IS A POTENTIAL
 CC TARGET FOR OXIDATIVE STRESS DURING CATARACT FORMATION IN LENS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN LENS AND RETINA (FOUND IN THE
 CC MUELLER CELLS). MODERATELY ABUNDANT IN HEART AND TESTIS (FOUND IN
 CC THE SERTOLI CELLS). AND PRESENT IN VERY LOW AMOUNTS IN LUNG.
 CC -1- SIMILARITY: BELONGS TO THE FABP/P2/CRBP/CRAPP FAMILY OF
 CC TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U55188; AAB41297.1; -;
 DR EMBL; AF059507; AAC14711.1; -;
 DR HSSP; Q01469; IBS6.
 DR InterPro; IPR000463; Fatty_acid_BP.
 DR InterPro; IPR000566; Lipocin_cytFABP.
 DR Pfam; PF00061; lipocalin; 1.
 DR PRINTS; PR00178; FATTYACIDBP.
 DR PROSITE; PS00214; FABP; 1.
 KW Transport; Lipid-binding; Phosphorylation.
 FT MOD_RES 22 22 PHOSPHORYLATION (BY TYR-KINASES)
 FT (PROBABLE)
 FT CONFLICT 52 52 L->P (IN REF. 1).
 FT SEQUENCE 135 AA; 15074 MW; 439B86AF68A34E2A CRC64;
 SQ
 Query Match 39.8%; Score 41; DB 1; Length 135;
 Best Local Similarity 47.6%; Pred. No. 9.9;
 Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;
 QY 1 NTTGTGVLKQWDG--STIS 19
 Db 88 NETDGLVQHOEWGDKSTIT 108
 RESULT 7
 ID FASE_HUMAN
 AC Q01469;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Fatty acid-binding protein, epidermal (E-FABP) (Psoriasis-associated
 DE fatty acid-binding protein homolog) (PA-FABP).

GN FAPB5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=92381332; PubMed=1512466;
 RA Madsen P.S., Rasmussen H.H., Leffers H., Honore B., Celis J.E.;
 RT "Molecular cloning and expression of a novel keratinocyte protein
 RT (psoriasis-associated fatty acid-binding protein [PA-FABP]) that is
 RT highly up-regulated in psoriatic skin and that shares similarity to
 RT fatty acid-binding proteins.";
 RL J. Invest. Dermatol. 99:299-305(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Utsdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Baha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Rosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 25-33; 39-50; 62-71; 83-101 AND 120-129.
 RC TISSUE=Keratinocytes;
 RX MEDLINE=93162043; PubMed=1286667;
 RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,
 RA Vandekerckhove J.;
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel
 RT protein database of normal human epidermal keratinocytes.";
 RL Electrophoresis 13:960-969(1992).
 RN [4]
 RP SEQUENCE OF 67-72 AND 104-110, AND CHARACTERIZATION.
 RX MEDLINE=94379963; PubMed=8092987;
 RA Sieganthal G., Hotz R., Chatellard-Gruaz D., Didierjean L.,
 RA Hallen U., Saurat J.-H.;
 RT "Purification and characterization of the human epidermal fatty acid-
 RT binding protein: localization during epidermal cell differentiation
 RT in vivo and in vitro.";
 RL Biochem. J. 302:363-371(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS).
 RX MEDLINE=99425110; PubMed=10493790;
 RA Hohoff C., Borchers T., Rustow B., Spener F., van Tilbeurgh H.;
 RT "Expression, purification and crystal structure determination of
 RT recombinant human epidermal-type fatty acid-binding protein.";
 RL Biochemistry 38:12229-12239(1999).
 CC -1- FUNCTION: HIGH SPECIFICITY FOR FATTY ACIDS. HIGHEST AFFINITY FOR
 CC C18 CHAIN LENGTH. DECREASING THE CHAIN LENGTH OR INTRODUCING
 CC DOUBLE BONDS REDUCES THE AFFINITY. MAY BE INVOLVED IN KERATINOCYTE
 CC DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: KERATINOCYTES; HIGHLY EXPRESSED IN PSORIATIC
 CC SKIN.
 CC -1- SIMILARITY: BELONGS TO THE FABP/P2/CRBP/CRAPP FAMILY OF
 CC TRANSPORTERS.

```
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M94856; AAA58467.1; -.
CC EMBL: BC019385; AAH19385.1; -.
CC PIR: I56326; I56326.
CC PDB: 1B56; 05-OCT-99.
CC PDB: 1JJJ; 03-JUL-02.
CC DR Aarhus/Ghent-2DPAGE; 3007; IEF.
CC DR Genew; HGNC:3560; FARP5.
CC MIM: 605168; -.
CC GO: GO:0005504; F: fatty acid binding activity; TAS.
CC GO: GO:0006544; P: epidermal differentiation; TAS.
CC GO: GO:0006629; P: lipid metabolism; TAS.
CC InterPro: IPR000463; Fatty_acid_BP.
CC InterPro: IPR000566; Lipocalin_cytFABP.
CC Pfam: PF00061; Lipocalin; 1.
CC PRINTS: PR00178; FATTYACIDBP.
CC PROSITE: PS00214; FARP; 1.
CC Transort; Lipid-binding; 3D-structure.
KW HELIX 4 7
FT STRAND 9 15
FT STRAND 19 26
FT TURN 27 27
FT HELIX 30 38
FT STRAND 42 48
FT TURN 49 50
FT STRAND 51 57
FT STRAND 62 68
FT TURN 69 70
FT STRAND 73 76
FT TURN 78 79
FT STRAND 82 90
FT TURN 91 92
FT STRAND 93 100
FT TURN 101 102
FT STRAND 103 112
FT TURN 113 114
FT STRAND 115 122
FT TURN 123 124
FT STRAND 125 133
FT SEQUENCE 135 AA; 15164 MW; 77D38F806143D63 CRC64;
Query Match 39.8%; Score 41; DB 1; Length 135;
Best Local Similarity 47.6%; Pred. No. 9,9;
Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;
QY 1 NNTTGVFLGKQDWG--STIS 19
DB 88 NFDGALVQHOEWGDKESTIT 108
RESULT 8
SYGB_PASMU STANDARD; PRT; 689 AA.
AC P57905;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE beta chain) (GlyRS).
GN GLYS OR PM1102.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustella M.L., Whittam T.S., Kapur V.;
RA "Complete genomic sequence of Pasteurella multocida Pm70.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
RL -!- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyl-tRNA(Gly).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006151; AAK03186.1; -.
CC HAMAP: MF_00255; 1.
CC InterPro: IPR002311; tRNA_synt_2f.
CC InterPro: IPR006194; tRNA_synt_Gly.
CC Pfam: PF02092; tRNA_synt_2f; 1.
CC PRINTS: PR01045; TRNASYNTHGB.
CC TIGFAMS; TIGR00211; glyS; 1.
CC PROSITE; PS50861; AA_TRNA_LIGASE_II_GLYAB; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 689 AA; 76019 MW; D67A980B5143B21E CRC64;
Query Match 39.8%; Score 41; DB 1; Length 689;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 NNTTGVFLGKQDWGGS 16
DB 459 DLTGIFGIGQPKGS 474
RESULT 9
YCBS_ECOLI
ID YCBS_ECOLI STANDARD; PRT; 866 AA.
AC P75857;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane usher protein ycbS precursor.
GN YCBS OR B0940.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
```

```

RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA RES. 3:137-155(1996).
CC -!- FUNCTION: COULD BE INVOLVED IN THE EXPORT AND ASSEMBLY OF THE
CC PUTATIVE YCBQ FIMBRIAL SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000196; AAC74026.1; -
DR EMBL; D90732; BAA35695.1; -
DR PIR; C64834; C64834.
DR EcoGene; EG13711; ycbS.
DR InterPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW Hypothetical protein; Outer membrane; Transmembrane; Fimbria;
KW Transport; Signal; Complete proteome.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 866 HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN
FT YCBS.
FT SQ SEQUENCE 866 AA; 95241 MW; 0004DC5E9FF1F5796 CRC64;
Query Match 39.8%; Score 41; DB 1; Length 866;
Best Local Similarity 63.6%; Pred. No. 72;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 8 GLKQDWGSTI 18
DB ||| ||| |
723 GLKTDWGRYTV 733
RESULT 10
GC5H_CHLPN STANDARD; PRT; 115 AA.
AC Q928B0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable glycine cleavage system H protein.
GN GC5H OR GC5H OR CPN0433 OR CP0320.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RA SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalnan S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
```

```

RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA".
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: THE GLYCINE CLEAVAGE SYSTEM CATALYZES THE DEGRADATION OF
CC GLYCINE. THE H PROTEIN SHUTTLES THE METHYLAMINE GROUP OF GLYCINE
CC FROM THE H PROTEIN TO THE T PROTEIN (BY SIMILARITY).
CC -!- COFACTOR: THE H CHAIN CONTAINS A COVALENTLY-BOUND LIPOYL COFACTOR
CC (BY SIMILARITY).
CC -!- SUBUNIT: The glycine cleavage system is composed of four proteins:
CC P, T, L, and H (By similarity).
CC -!- SIMILARITY: Belongs to the gcvH family.
CC -!- SIMILARITY: Contains 1 lipoyl-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001626; AAD18577.1; -
DR EMBL; AE002194; AAF38175.1; -
DR EMBL; AF002346; BAA98641.1; -
DR PIR; C72079; C72079.
DR PIR; G86544; G86544.
DR HSSP; PI6048; LHTP.
DR TIGR; CP0320; -.
DR HAMAP; MF_00272; -.
DR InterPro; IPR002930; GCV_H.
DR InterPro; IPR003016; LipoyL.
DR Pfam; PF01597; GCV_H; 1.
DR PROSITE; PS00189; LIPOYL; FALSE_NEG.
KW LipoyL; Complete proteome.
FT BINDING 58 58 LIPOYL (BY SIMILARITY).
FT SQ SEQUENCE 115 AA; 12876 MW; 43FF4363A48D175 CRC64;
Query Match 38.8%; Score 40; DB 1; Length 115;
Best Local Similarity 57.1%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 6 VFGLKQDWGSTIS 19
DB ||| ||| |
97 VVRLLQDWDFSNLS 110
RESULT 11
RBS_MARPA
ID RBS_MARPA STANDARD; PRT; 180 AA.
AC O64416;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ribulose biphosphate carboxylase small chain, chloroplast precursor
DE (EC 4.1.1.39) (RuBisCO small subunit).
GN RBS.
OS Marchantia paleacea.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
OC Marchantiopsida; Marchantiales; Marchantiales; Marchantiaceae;
OC Marchantiaceae; Marchantia.
OX NCBI_TaxID=56867;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Var. diptera; TISSUE=Callus;
RX MEDLINE=99205705; PubMed=10189707;
RA Suzuki T., Takio S., Tanaka K., Yamamoto I., Satoh T.;
```

RT "Differential light regulation of the rbcS gene expression in two cell lines of the liverwort *Marchantia paleacea* var. *diptera*.";

RL Plant Cell Physiol. 40:100-103(1999).

CC -!- FUNCTION: Rubisco catalyzes two reactions: the carboxylation of D-ribulose 1,5-bisphosphate, the primary event in photosynthetic carbon dioxide fixation, as well as the oxidative fragmentation of the pentose substrate in the photorespiration process. Both reactions occur simultaneously and in competition at the same active site.

CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + CO(2) = 2 3-phospho-D-glycerate.

CC -!- CATALYTIC ACTIVITY: D-ribulose 1,5-bisphosphate + O(2) = 3-phospho-D-glycerate + 2-phosphoglycolate.

CC -!- SUBUNIT: 8 large chains + 8 small chains.

CC -!- SUBCELLULAR LOCATION: Chloroplast.

CC -!- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB004883; BAA28610.1; -.

DR InterPro; IPR000894; RuBisCO_small.

DR Pfam; PF00101; RUBISCO_small; 1.

DR PRINTS; PR00152; RUBISCO_SMALL.

DR ProDom; PD000290; RUBISCO_small; 1.

KW Photosynthesis; Carbon dioxide fixation; Photorespiration; Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide.

FT TRANSIT 1 58 CHLOROPLAST (BY SIMILARITY).

FT CHAIN 59 180 RUBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN.

SQ SEQUENCE 180 AA; 19869 MW; D736A467732DCB4 CRC64;

Query Match 38.8%; Score 40; DB 1; Length 180;

Best Local Similarity 42.9%; Pred. No. 20;

Matches 9; Conservative 3; Mismatches 3; Indels 6; Gaps 1;

OY 5 GVFGKQDW-----DGTST 19
||||| :| :| :| :|

DB 37 GVFGKSEWQTKTQNGSRVS 57

RESULT 12

BIOS_HELPJ

ID BIOS_HELPJ STANDARD; PRT; 282 AA.

AC Q9JTK8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).

GN Btob OR JHP1298.

OS Helicobacter pylori J99 (Campylobacter pylori J99).

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.

OX NCBI_TaxID=85963;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C., Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G., Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C., Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F., Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*.";

RL Nature 397:176-180(1999).

CC -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.

CC -!- PATHWAY: Biotin biosynthesis; last step.

CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES

FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF001553; AAD06876.1; -.

DR PIR; H71823; H71823.

DR InterPro; IPR002884; Biotin_synth.

DR InterPro; IPR006638; Eip3.

DR Pfam; PF04055; Radical_SAM; 1.

DR SMART; SM00729; Eip3; 1.

DR TIGRFAMs; TIGR00433; biob; 1.

KW Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.

FT METAL 17 17 IRON-SULFUR (POTENTIAL).

FT METAL 21 21 IRON-SULFUR (POTENTIAL).

FT METAL 24 24 IRON-SULFUR (POTENTIAL).

SQ SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAF1B CRC64;

Query Match 38.8%; Score 40; DB 1; Length 282;

Best Local Similarity 50.0%; Pred. No. 32;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 GVFGKQDW 14
||||| :| :| :|

DB 157 GIFGLNESWE 166

RESULT 13

BIOS_HELPJ

ID BIOS_HELPJ STANDARD; PRT; 282 AA.

AC O25956;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).

GN Btob OR HPI1406.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.

OX NCBI_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=26695 / ATCC 700392;

RA MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G., Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A., Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S., Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A., McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K., Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M., Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E., Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M., Venter J.C.;

RT "The complete genome sequence of the gastric pathogen *Helicobacter pylori*.";

RL Nature 388:539-547(1997).

CC -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.

CC -!- PATHWAY: Biotin biosynthesis; last step.

CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Glutamyl-tRNA synthetase 1 (EC 6.1.1.17) (Glutamate--tRNA ligase 1)
 DE (Glurs 1)
 GN GLTX1 OR BR1147.
 OS Brucella suis.
 OC Bacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29461;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=1330 / Biovar 1;
 RX MEDLINE=2247741; PubMed=12271122;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
 RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
 RA Riedmuller S., Tettelin H., Gill S.K., White O., Salzberg S.L.,
 RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
 CC diphosphate + L-glutamyl-tRNA(Glu).
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF014415; AAN30067.1; -.
 DR TIGR; BR1147; -.
 DR HAMAP; MF_00022; -; 1.
 DR InterPro; IPR004527; Gltx_bact.
 DR InterPro; IPR000924; Glu_tRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR PRINTS; TIGR00987; TRNASYNTHGLU.
 DR TIGRFAMS; TIGR00464; gltx_bact; 1.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT Site 11 21 "HIGH" REGION.
 FT SITE 240 244 "KMSKS" REGION.
 FT BINDING 243 243 ATP (BY SIMILARITY).
 SQ SEQUENCE 473 AA; 51981 MW; CDDF287970001020 CRC64;
 Query Match 38.8%; Score 40; DB 1; Length 473;
 Best Local Similarity 53.3%; Pred. No. 55;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 5 GVFGKQWDGSGTIS 19
 I: | ||||| I
 Db 62 GLTWGLDWDGEAIS 76
 RESULT 17
 MTHC DROME
 ID MTHC DROME STANDARD; PRT; 475 AA.
 AC P83119;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable G-protein-coupled receptor Mth-like 12 precursor (Methuselah-
 DE like 12 protein).
 GN MTHL12 OR MTH-LIKE-12.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aqbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Fodor C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoston K., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shieh H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spiers E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 RN IDENTIFICATION.
 RP MEDLINE=21173629; PubMed=11274391;
 RX West A.P. Jr., Llamas L.L., Snow P.M., Benzer S., Bjorkman P.J.;
 RA "Crystal structure of the ectodomain of Methuselah, a Drosophila G
 RT protein-coupled receptor associated with extended lifespan."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3744-3749(2001).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED
 CC RECEPTORS. MTH SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF003699; -; NOT_ANNOTATED_CDS.
 DR FlyBase; FBgn0045442; mthl12.
 DR GO; GO:0004930; P:G-protein coupled receptor activity; ISS.
 DR GO; GO:0008340; P:determination of adult life span; ISS.
 DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; ISS.
 DR GO; GO:0006950; P:response to stress; ISS.
 DR InterPro; IPR000832; GPCR_secretin.

```

DR Pfam: PF00002; 7tm_2; 1.
DR PROSITE; PS00649; G_PROTEIN_RECP_F2.1; FALSE_NEG.
DR PROSITE; PS00650; G_PROTEIN_RECP_F2.2; FALSE_NEG.
DR PROSITE; PS0261; G_PROTEIN_RECP_F2.4; 1.
KW Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Signal; Multigene family.
FT FT CHAIN 1 17
FT FT CHAIN 18 475
FT FT DOMAIN 18 214
FT FT TRANSMEM 215 235
FT FT DOMAIN 236 242
FT FT TRANSMEM 243 263
FT FT DOMAIN 264 275
FT FT TRANSMEM 276 296
FT FT DOMAIN 297 307
FT FT TRANSMEM 308 328
FT FT DOMAIN 329 360
FT FT TRANSMEM 361 381
FT FT DOMAIN 382 403
FT FT TRANSMEM 404 424
FT FT DOMAIN 425 442
FT FT TRANSMEM 443 463
FT FT DOMAIN 464 475
FT FT DISULFID 27 81
FT FT DISULFID 83 88
FT FT DISULFID 92 183
FT FT DISULFID 93 104
FT FT DISULFID 149 203
FT FT CARBOHYD 19 19
FT FT CARBOHYD 34 34
FT FT CARBOHYD 55 55
FT FT CARBOHYD 135 135
FT FT CARBOHYD 352 352
FT FT CARBOHYD 475 AA; 55424 MW; 3590FEF875F77264 CRC64;
SQ SEQUENCE 475 AA; 55424 MW; 3590FEF875F77264 CRC64;

Query Match 38.8%; Score 40; DB 1; Length 475;
Best Local Similarity 53.8%; Pred. No. 55;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 GYFGLQKQWDGST 17
Db 343 GYFGLFTDWNSS 355

RESULT 18
PLOC_MOUSE
AC Q9R0E1; Q9CY9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 precursor
DE (EC 1.14.11.4) (Lysyl hydroxylase 3) (LH3).
GN PLOC3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=99357020; PubMed=10429951;
RA Ruotsalainen H., Sipila L., Kerkela E., Pospiech H., Myllylae R.;
RT "Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3,
RT their phylogenetic analysis and tissue-specific expression in the
RT mouse.";
RL Matrix Biol. 18:325-329(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=21233587; PubMed=11334715;
RA Ruotsalainen H., Vanhatupa S., Tampio M., Sipila L., Vaitavaara M.,
RA Myllylae R.;

```

```

RT RT *Complete genomic structure of mouse lysyl hydroxylase 2 and lysyl
RL RL hydroxylase 3/collagen glucosyltransferase.*;
RN RN Matrix Biol. 20:137-146(2001).
RP [3]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
RA Gustincich S., Hill D., Hofmann M., Fujita M., Gariboldi M.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Willing L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
CC -1- FUNCTION: FORMS HYDROXYLYSINE RESIDUES IN -XAA-LYS-GLY- SEQUENCES
CC IN COLLAGENS. THESE HYDROXYLYSINES SERVE AS SITES OF ATTACHMENT
CC FOR CARBOHYDRATE UNITS AND ARE ESSENTIAL FOR THE STABILITY OF THE
CC INTERMOLECULAR COLLAGEN CROSSLINKS.
CC -1- CATALYTIC ACTIVITY: Procollagen L-lysine + 2-oxoglutarate + O(2) =
CC Procollagen 5-hydroxy L-lysine + succinate + CO(2).
CC -1- PROCOFFACTOR: IRON AND ASCORBATE.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: MEMBRANE BOUND IN CISTERNAE OF ROUGH
CC ENDOPLASMIC RETICULUM.
CC -1- TISSUE SPECIFICITY: Highly expressed in the heart, lung, liver and
CC testis.
CC -1- SIMILARITY: BELONGS TO THE LYSYL HYDROXYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF046783; AAD54618.1; -
CC EMBL: AY014830; AAK00576.1; -
CC EMBL: AK013195; BAB28704.1; -
CC MGD: MGI:1347008; Plod3
CC DR InterPro: IPR005123; 2OG-FeII_Oxy
CC DR InterPro: IPR006620; PRO_4_hyd_alph.
CC DR Pfam: PF03171; 2OG-FeII_Oxy.1.
CC DR Pfam: PF011578; Procollys_dioxy; 1.
CC DR SMART: SM00702; P4HC; 1.
CC DR PROSITE; PS01325; LYS_HYDROXYLASE; 1.
CC DR Oxidoreductase; Dioxygenase; Signal; Iron; Vitamin C; Glycoprotein;
KW Endoplasmic reticulum; Membrane; Polymorphism.
FT SIGNAL 1 27
FT CHAIN 28 741
FT FT METAL 670 670
FT FT METAL 672 672
FT FT METAL 722 722
FT FT ACT_SITE 732 732
FT FT CARBOHYD 66 66
FT FT CARBOHYD 286 286
FT FT CARBOHYD 551 551
FT FT CONFLICT 8 8

```

```
SQ SEQUENCE 741 AA; 84922 MW; D1B79B386339D9F4 CRC64;
Query Match 38.8%; Score 40; DB 1; Length 741;
Best Local Similarity 36.8%; Pred. No. 89;
Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 NNTTGVFLKQDWDGSGTIS 19
| | | | | | | | | |
66 NYTVRTLGLGQEWGRGDVA 84

RESULT 19
ENV_CAEVG
ID ENV_CAEVG STANDARD; PRT; 942 AA.
AC P31627;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Env polyprotein precursor (Coat polyprotein) [Contains: Surface
DE protein; Transmembrane protein].
GN ENV.
OS Caprine arthritis encephalitis virus (strain G63) (CAEV).
OC Viruses; Retroviral viruses; Retroviridae; Lentiviruses.
OX NCBI_TaxID=11562;
RN [1]
RP SEQUENCE FROM N.A.; AND SEQUENCE OF 81-95.
RX MEDLINE=92015464; PubMed=1656067;
RA Knowles D.P. Jr., Cheevers W.P., McGuire T.C., Brassfield A.L.,
RA Harwood W.G., Stem T.A.;
RT "Structure and genetic variability of envelope glycoproteins of two
RT antigenic variants of caprine arthritis-encephalitis lentivirus.";
RT J. Virol. 65:5744-5750(1991).
RN [2]
RP REVISIONS.
RA Knowles D.P.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M60855; AAB88709.2; -.
DR InterPro: IPR000328; Env_GP41.
DR Pfam: PF00517; GP41; 1.
KW Glycoprotein; Coat protein; Polypeptide; Transmembrane.
FT PEPTIDE 1 80
FT CHAIN 81 630
FT CHAIN 631 942
FT DOMAIN 1 630
FT TRANSMEM 631 659
FT DOMAIN 660 799
FT TRANSMEM 800 820
FT DOMAIN 821 942
FT CARBOHYD 51 51
FT CARBOHYD 98 98
FT CARBOHYD 131 131
FT CARBOHYD 176 176
FT CARBOHYD 228 228
FT CARBOHYD 331 331
FT CARBOHYD 348 348
FT CARBOHYD 354 354
FT CARBOHYD 370 370
FT CARBOHYD 379 379
FT CARBOHYD 400 400
FT CARBOHYD 404 404
FT CARBOHYD 435 435
FT CARBOHYD 441 441
FT CARBOHYD 447 447
FT CARBOHYD 457 457
```

```
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 503 503 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 942 AA; 108437 MW; 5B0E1F0F3D355F4A CRC64;

Query Match 38.8%; Score 40; DB 1; Length 942;
Best Local Similarity 46.2%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 NNTTGVFLKQDQW 13
| | | | | | | |
482 NTITGIMGTNTNW 494

RESULT 20
DPYD_HUMAN
ID DPYD_HUMAN STANDARD; PRT; 1025 AA.
AC Q12882; Q16694; Q16761;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dihydropyrimidine dehydrogenase [NADP+] precursor (EC 1.3.1.2) (DPD)
DE (DHPDHase) (Dihydrouracil dehydrogenase) (Dihydrothymine
DE dehydrogenase).
GN DPYD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94365020; PubMed=8083224;
RA Yokota H., Fernandez-Salguero P., Furuya H., Lin K., McBride O.W.,
RA Podschun B., Schnackerz K.D., Gonzalez F.J.;
RT "cDNA cloning and chromosome mapping of human dihydropyrimidine
RT dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and
RT congenital thymine uraciluria.";
RT J. Biol. Chem. 269:23192-23196(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97280676; PubMed=9135003;
RA Johnson M.R., Wang K., Tillmanns S., Albin N., Diasio R.B.;
RT "Structural organization of the human dihydropyrimidine dehydrogenase
RT gene.";
RT Cancer Res. 57:1660-1663(1997).
RN [3]
RP SEQUENCE OF 581-635 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97047101; PubMed=8892022;
RA Vreken P., van Kuilenburg A.B.P., Meinsma R., Smit G.P.A.,
RA Bakker H.D., de Abreu R.A., van Gennip A.H.;
RT "A point mutation in an invariant splice donor site leads to exon
RT skipping in two unrelated Dutch patients with dihydropyrimidine
RT dehydrogenase deficiency.";
RT J. Inherit. Metab. Dis. 19:645-654(1996).
RN [4]
RP CHARACTERIZATION, AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=92381021; PubMed=1512248;
RA Lu Z.-H., Zhang R., Diasio R.B.;
RT "Purification and characterization of dihydropyrimidine dehydrogenase
RT from human liver.";
RT J. Biol. Chem. 267:17102-17109(1992).
RN [5]
RP VARIANTS ARG-29; TRP-235 AND HIS-886.
RX MEDLINE=98102836; PubMed=9439663;
RA Vreken P., van Kuilenburg A.B.P., Meinsma R., van Gennip A.H.;
RT "Dihydropyrimidine dehydrogenase (DPD) deficiency: identification and
```



```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U05958; AAB60617.1; -.
CC DR PIR; A53491; A53491.
CC DR InterPro; IPR002293; AA/rel_permeasel.
CC DR InterPro; IPR004842; KCL_cotranspt.
CC DR PRINTS; PR01207; NAKCL_cotranspt.
CC DR TIGR; TIGR00930; 2a30; 1.
CC KW Transport; Ion transport; Sodium transport; Symport;
CC Potassium transport; Potassium; Transmembrane; Phosphorylation.
CC FT DOMAIN 1 237 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 258 278 POTENTIAL.
CC FT TRANSMEM 282 302 POTENTIAL.
CC FT DOMAIN 303 339 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 340 360 POTENTIAL.
CC FT TRANSMEM 383 403 POTENTIAL.
CC FT DOMAIN 404 407 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 408 428 POTENTIAL.
CC FT TRANSMEM 459 479 POTENTIAL.
CC FT DOMAIN 480 496 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 497 517 POTENTIAL.
CC FT TRANSMEM 565 585 POTENTIAL.
CC FT DOMAIN 586 630 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 631 651 POTENTIAL.
CC FT TRANSMEM 652 672 POTENTIAL.
CC FT DOMAIN 673 689 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 690 710 POTENTIAL.
CC FT TRANSMEM 776 796 POTENTIAL.
CC FT DOMAIN 797 1191 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 27 30 POLY-PRO.
CC FT DOMAIN 51 57 POLY-GLY.
CC FT DOMAIN 67 70 POLY-ALA.
CC FT DOMAIN 89 92 POLY-ALA.
CC FT DOMAIN 96 99 POLY-ALA.
CC FT DOMAIN 126 129 POLY-PRO.
CC FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT MOD_RES 189 189 PHOSPHORYLATION.
CC FT MOD_RES 1114 1114 PHOSPHORYLATION.
CC SQ SEQUENCE 1191 AA; 129774 MW; BA7BB9815431500C CRC64;

Query Match 38.8%; Score 40; DB 1; Length 1191;
Best Local Similarity 70.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 TGVEGLKQDW 13
DB 856 TLVFGFKDW 865

RESULT 22
GCST_NEIMB
ID GCST_NEIMB STANDARD; PRT; 366 AA.
AC Q9K0L8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable aminomethyltransferase (EC 2.1.2.10) (Glycine cleavage system
DE T protein).
GN GCVT OR NMB0574.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,

```

```

RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittiore H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of
CC glycine (By similarity).
CC -!- CATALYTIC ACTIVITY: (6S)-tetrahydrofolate + S-
CC aminomethyldihydropyrimidinol + NH(3) -> (6R)-5,10-
CC methylenetetrahydrofolate + NH(3) + dihydropyrimidinol.
CC -!- SUBUNIT: The glycine cleavage system is composed of four proteins:
CC P, T, L, and H (By similarity).
CC -!- SIMILARITY: Belongs to the gcvT family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE002413; AAF41002.1; -.
CC DR PIR; A81183; A81183.
CC DR TIGR; NMB0574; -.
CC DR HAMAP; MF_00259; -.
CC DR InterPro; IPR006222; GCV_T.
CC DR InterPro; IPR006223; GcvT.
CC DR Pfam; PF01571; GCV_T; 1.
CC DR TIGRFAMS; TIGR00528; gcvT; 1.
CC DR Transferase; Aminotransferase; Complete proteome.
CC SQ SEQUENCE 366 AA; 39740 MW; 566A7BCD21691D91 CRC64;

Query Match 38.3%; Score 39.5; DB 1; Length 366;
Best Local Similarity 44.4%; Pred. No. 51;
Matches 12; Conservative 2; Mismatches 2; Indels 11; Gaps 2;

QY 2 TTTGVF--GLKQ-----DWDGST 17
DB 312 TTTGVFSPSLKQSIATIAIAPKDFDGD 338

RESULT 23
GCST_NEIMA
ID GCST_NEIMA STANDARD; PRT; 368 AA.
AC Q9JVP2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable aminomethyltransferase (EC 2.1.2.10) (Glycine cleavage system
DE T protein).
GN GCVT OR NMA0758.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of

```

CC glycine (By similarity).

CC -1- CATALYTIC ACTIVITY: (6S)-tetrahydrofolate + S-

CC aminomethylidihydrofolate + NH(3) + dihydrodipolypyrroline.

CC -1- SUBUNIT: The glycine cleavage system is composed of four proteins:

CC P, T, L, and H (By similarity).

CC -1- SIMILARITY: Belongs to the gcvT family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL: AL162754; CAB84041.1; -

CC PIR: G81919; G81919.1

CC HAMAP: MF_00259; 1

CC InterPro: IPR006222; GCV_T.

CC Pfam: PF01571; GCV_T; 1.

CC TIGRFAMs: TIGR00528; gcvT; 1.

CC KW Transferase; Aminotransferase; Complete proteome.

CC SQ SEQUENCE 368 AA; 39910 MW; C50AB7FA924CIABF CRC64;

Query Match 38.3%; Score 39.5; DB 1; Length 368;

Best Local Similarity 44.4%; Pred. No. 51;

Matches 12; Conservative 2; Mismatches 2; Indels 11; Gaps 2;

QY 2 TTGTGVP-GLKQ-----DWDGST 17

II:III III I:III I

Db 314 TTSGVSPSLKOSIAIARVPKDFDGD 340

RESULT 24

DTD_CLOB

ID DTD_CLOB STANDARD; PRT: 149 AA.

AC Q37G02; 2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE D-TYROSYL-CRNA(Tyr) deacylase (EC 3.1.-.-).

GN DTD OR CAC2273.

OS Clostridium acetobutylicum.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

OC Clostridium.

OX NCBI_TaxID=1488;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;

RX MEDLINE=21359325; PubMed=11466286;

RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,

RA Gibson R., Lee H.M., Dubois J., Olu D., Hitti J., Wolf Y.I.,

RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,

RA Bennett G.N., Koonin E.V., Smith D.R.;

RA "Genome sequence and comparative analysis of the solvent-producing

RT bacterium Clostridium acetobutylicum.";

RL J. Bacteriol. 183:4823-4838(2001).

CC -1- FUNCTION: Hydrolyzes D-tyrosyl-tRNA(Tyr) into D-tyrosine and free

CC of D-tyrosine (By similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- SIMILARITY: BELONGS TO THE DTD FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

DR EMBL: AB007728; AAK80230.1; -

DR PIR: C97180; C97180.

DR HSSP: P32147; 1JKE.

DR HAMAP: MF_00518; -; 1.

DR InterPro: IPR003732; DTYrRNA_deacyls.

DR Pfam: PF02580; Tyr_Deacylase; 1.

DR ProDom: PD005653; DTYrRNA_deacyls; 1.

DR TIGRFAMs: TIGR00256; TIGR00256; 1.

DR KW Hydrolase; Complete proteome.

CC

CC Query Match 37.9%; Score 39; DB 1; Length 149;

CC Best Local Similarity 57.1%; Pred. No. 23;

CC Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWD 14

II:III I:III I

Db 119 NVQTGVFGAHMDVD 132

RESULT 25

EFPL_XYLF

ID EFPL_XYLF STANDARD; PRT: 188 AA.

AC Q87C43;

DT 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Elongation factor P-like protein.

GN PD1253.

OS Xylella fastidiosa (strain Temeculal / ATCC 700964).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xylella.

OX NCBI_TaxID=183190;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22421331; PubMed=12533478;

RA Van Siuys M.A., de Oliveira M.C., Monteiro-Vitorello C.B., Moon D.H.,

RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., da Silva F.R.,

RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,

RA Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,

RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,

RA Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,

RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Anaral A.M.,

RA Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,

RA da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,

RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,

RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,

RA Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,

RA Kitajima J.P.;

RT "Comparative analyses of the complete genome sequences of Pierce's

RT disease and citrus variegated chlorosis strains of Xylella

RT fastidiosa";

RL J. Bacteriol. 185:1018-1026(2003).

CC -1- SIMILARITY: Belongs to the elongation factor P family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC

CC EMBL: AB012557; AAO29102.1; -

CC HAMAP: MF_00646; -; 1.

CC PROSITE: PS01275; EFP; 1.

CC Complete proteome.

CC

CC Query Match 37.9%; Score 39; DB 1; Length 188;

CC Best Local Similarity 87.5%; Pred. No. 30;

CC Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY SEQUENCE 188 AA; 20444 MW; 114A94851DCD5514 CRC64;

```
QY 1 NNTTGVFG 8
Db 177 NNTTGAFG 184

RESULT 26
EFPL_XYLFA
ID EFPL_XYLFA STANDARD; PRT; 189 AA.
AC Q9PBE1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Elongation factor P-like protein.
GN XF2203.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lopes M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RA "The genome sequence of the plant pathogen Xylella fastidiosa.";
RT Nature 406:151-159(2000).
RL -!- SIMILARITY: Belongs to the elongation factor P family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE004033; AAF05002.1; -
CC HAMAP: MF_00646; -; 1.
CC InterPro: IPR001059; EF-P.
CC Pfam: PF01132; EFP; 1.
CC PROSITE: PS01275; EFP; 1.
CC Complete proteome.
KW SEQUENCE 189 AA; 20616 MW; C809AC6DF0A51A5B CRC64;

Query Match 37.9%; Score 39; DB 1; Length 189;
Best Local Similarity 87.5%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNTTGVFG 8
Db 178 NNTTGAFG 185

RESULT 27
PUNA_CELSP
ID PUNA_CELSP STANDARD; PRT; 282 AA.
AC P81989;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase)
DE (PNP).
GN PUNA.
OS Cellulomonas sp.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococccineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=40001;
RN [1]
RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=20069945; PubMed=10600382;
RA Tebbe J., Bzowska A., Wielgus-Kutrowska B., Schroeder W.,
RA Kazmierczuk Z., Shugar D., Saenger W., Koellner G.;
RT "Crystal structure of the purine nucleoside phosphorylase (PNP) from
RT Cellulomonas sp. and its implication for the mechanism of trimeric
RL PNPs.";
RL J. Mol. Biol. 294:1239-1255(1999).
CC -!- FUNCTION: CLEAVAGE OF GUANOSINE OR INOSINE TO RESPECTIVE BASES AND
CC SUGAR-1-PHOSPHATE MOLECULES.
CC -!- CATALYTIC ACTIVITY: Purine nucleoside + phosphate = purine +
CC alpha-D-ribose 1-phosphate.
CC -!- PATHWAY: Purine nucleoside salvage.
CC -!- SUBUNIT: Homotrimer.
CC -!- SIMILARITY: BELONGS TO THE PNP/MTAP FAMILY 2 OF PHOSPHORYLASES.
CC PDB: 1C3X; 29-JUL-99.
CC PDB: 1QF5; 12-DEC-99.
CC InterPro: IPR001369; Mtap_PNP.
CC Pfam: PF00896; Mtap_PNP; 1.
CC PROSITE: PS01240; PNP_MTPAP.2; FALSE_NEG.
KW Transferase; Glycosyltransferase; 3D-structure.
FT DOMAIN 1 204
FT ACT_SITE 204 204
FT TURN 11 12
FT TURN 14 15
FT HELIX 18 33
FT STRAND 40 43
FT TURN 46 47
FT TURN 49 54
FT STRAND 57 63
FT HELIX 64 66
FT TURN 68 69
FT STRAND 81 88
FT TURN 90 91
FT STRAND 94 99
FT HELIX 106 108
FT HELIX 112 115
FT HELIX 117 124
FT TURN 125 126
FT STRAND 129 138
FT TURN 141 142
FT TURN 145 146
FT STRAND 148 156
FT TURN 172 173
FT STRAND 175 175
FT HELIX 177 186
FT TURN 188 189
FT STRAND 192 197
FT HELIX 206 215
FT TURN 216 216
FT STRAND 219 221
FT HELIX 225 233
FT TURN 234 235
FT STRAND 237 247
```

```

FT TURN 249 251
FT HELIX 258 280
FT TURN 281 282
SQ SEQUENCE 282 AA; 29021 MW; 65F468DACCA3D360 CRC64;

Query Match 37.9%; Score 39; DB 1; Length 282;
Best Local Similarity 46.7%; Pred No. 46;
Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 TTGVFLGKQDWDGST 17
DB 133 TNGCGGLNQEWGAGT 147

RESULT 28
YQAK_BACSU
ID YQAK_BACSU STANDARD; PRT; 284 AA.
AC P45908;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yqak.
GN YQAK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
[1]
PC SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=95219086; PubMed=7704261;
RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;
RT "Complete nucleotide sequence of a skin element excised by DNA
rearrangement during sporulation in Bacillus subtilis.",
RL Microbiology 141:323-327(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RB Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
the Bacillus subtilis genome containing the skin element and many
sporulation genes.",
RN Microbiology 142:3103-3111(1996).
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brisset S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Fritz C.D., Fittington J., Fabret C., Ferrarri E., Foulger D.,
RA Fritz S.Y., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guilbert G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi K., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapkus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Leplante A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellouk R.F., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Plescan E., Pujic B., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Rochet B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

```

```

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.",
RL Nature 390:249-256(1997).
RN [4]
RP IDENTIFICATION.
RX MEDLINE=96084975; PubMed=7489895;
RA Medigue C., Moszer I., Viari A., Danchin A.;
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
computer system prototype.",
RL Gene 165:GC37-GC51(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D32216; BAA06925.1;
CC EMBL; D84432; BAA12386.1;
CC EMBL; Z99117; CAB14569.1;
CC PIR; B69945; B69945.
CC Subtilist; BG11262; YQAK.
CC InterPro; IPR004590; Rect.
CC Pfam; PF03837; Rect. 1.
CC TIGRFAMs; TIGR00616; rect. 1.
CC Hypothetical protein; Complete proteome.
SQ SEQUENCE 284 AA; 32170 MW; F255261D4692ADB7 CRC64;

Query Match 37.9%; Score 39; DB 1; Length 284;
Best Local Similarity 50.0%; Pred No. 47;
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 FGLKQDWDGSTI 18
DB 193 FGKNDWDAMAL 204

RESULT 29
SYGB_HAEAE
ID SYGB_HAEAE STANDARD; PRT; 285 AA.
AC O30836;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE beta chain) (GLYRS) (Fragment).
GN GLYS.
OS Haemophilus aegyptius.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=197575;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FJ031;
RA Tondella M.L.C.; Utt E.A., Mayer L.W., Quinn F.D.;
RA Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RL -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
+ glycyl-tRNA(Gly).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains
(RY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AF018635; AAB70306.1; -;
DR HAMAP; MF_00235; -; 1.
DR InterPro; IPR002311; trna_synt_2f.
DR InterPro; IPR006194; trna_synt_Gly.
DR Pfam; PF02092; trna_synt_2f; 1.
DR PROSITE; PS50861; AA_TRNA_LIGASE_II_GLYAB; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT NON_TER 1
SQ SEQUENCE 285 AA; 30957 MW; 034C4EE695DBB138 CRC64;

Query Match 37.9%; Score 39; DB 1; Length 285;
Best Local Similarity 50.0%; Pred. No. 47;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGDS 16
: |||: |||
Db 56 DTLTGIFGIGQAPKGS 71

RESULT 30

HEM3_FUSNN STANDARD; PRT; 298 AA.
ID HEM3_FUSNN
AC Q8RFP5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Porphobilinogen deaminase (EC 4.3.1.8) (PBG) (Hydroxymethylbilane
synthase) (HMB) (Pre-uroporphyrinogen synthase).
GN HEMC OR FN0645.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhatnagar A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
CC -!- FUNCTION: Tetrapolymerization of the monopyrrole PBG into the
hydroxymethylbilane preuroporphyrinogen in several discrete steps.
CC -!- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =
hydroxymethylbilane + 4 NH(3).
CC -!- COFACTOR: Covalently binds a dipyrromethane cofactor to which the
porphobilinogen subunits are added (by similarity).
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; fourth step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE HMB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AE010575; AAL94841.1; -;
DR HAMAP; MF_00260; -; 1.
DR InterPro; IPR000860; Porphobil_deam.
DR Pfam; PF01379; Porphobil_deam; 1.
DR Pfam; PF03900; Porphobil_deamC; 1.
DR ProDom; PD002745; Porphobil_deam; 1.
DR TIGRFAMs; TIGR00212; hencC; 1.

DR PROSITE; PS00533; PORPHOBILINOGEN_DEAM; 1.
KW Porphyrin biosynthesis; Lyase; Complete proteome.
FT BINDING 242 242 PYRROMETHANE COFACTOR (BY SIMILARITY).
SQ SEQUENCE 298 AA; 33098 MW; 13908A7DAA56984 CRC64;

Query Match 37.9%; Score 39; DB 1; Length 298;
Best Local Similarity 47.1%; Pred. No. 49;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 TTGVFLKQDWGDS 19
: |||: |||
Db 42 TSGDKDLKSNWNSDIS 58

RESULT 31

YM16_MYCLE STANDARD; PRT; 307 AA.
ID YM16_MYCLE
AC O32960;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ML0860.
GN ML0860 OR MLCB22.18.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: BELONGS TO THE UPF0105 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; Z98741; CAB11383.1; -;
DR EMBL; AL583920; CAC31241.1; -;
DR PIR; T44893; T44893.
DR Leproma; ML0860; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 307 AA; 32853 MW; EEP5BA102455DA2 CRC64;

Query Match 37.9%; Score 39; DB 1; Length 307;
Best Local Similarity 60.0%; Pred. No. 51;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 9 LKQDWGDS 18
: |||: |||
Db 143 LCQDWEGATL 152

RESULT 32

GBLP_NICPL STANDARD; PRT; 326 AA.
ID GBLP_NICPL
AC P93340;
DT 15-JUL-1998 (Rel. 36, Created)

```

15-JUL-1998 (Rel. 36, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit-like protein.
OS Nicotiana plumbaginifolia (Leadwort-leaved tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20225456; PubMed=10760577;
RA "Molecular characterization of cDNAs encoding G protein alpha and
RT beta subunits and study of their temporal and spatial expression
RL patterns in Nicotiana plumbaginifolia Viv.",
RL Biochim. Biophys. Acta 1491:143-160(2000).
CC -!- FUNCTION: MAY HAVE A ROLE IN CELL DIVISION.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; Y09514; CAA70705.1;
CC PIR; T16987; T16987.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 7.
CC ProDom; PD0000018; WD40; 4.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00678; WD_REPEATS_1; 4.
CC PROSITE; PS00082; WD_REPEATS_2; 6.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
CC REPEAT; WD repeat.
CC REPEAT 14 45 WD 1.
CC REPEAT 102 92 WD 2.
CC REPEAT 104 134 WD 3.
CC REPEAT 104 180 WD 4.
CC REPEAT 193 222 WD 5.
CC REPEAT 233 262 WD 6.
CC REPEAT 292 322 WD 7.
CC SEQUENCE 326 AA; 35820 MW; 92C80F769CAC8100 CRC64;
Query Match 37.9%; Score 39; DB 1; Length 326;
Best Local Similarity 42.9%; Pred. No. 54;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
OY 2 TTGTGVLKQDWG 15
Db ::|:|:|
73 SSDGMFALSGWDG 86
RESULT 33
GBLP_TOBAC STANDARD; PRT; 326 AA.
AC P49026;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit-like protein.
GN ARCA.
OS Nicotiana Tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Bright Yellow 2;
RX MEDLINE=94068562; PubMed=8248221;

```

```

RA "Isolation of cDNA of an auxin-regulated gene encoding a G protein
RT beta subunit-like protein from tobacco BY-2 cells.",
RL Proc. Natl. Acad. Sci. U.S.A. 90:11152-11156(1993).
CC -!- FUNCTION: MAY HAVE A ROLE IN CELL DIVISION.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN ROOTS. ALSO FOUND
CC IN LEAVES, FLOWER BUDS, SHOOT TIPS, STEMS AND YOUNG SEEDLINGS.
CC -!- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U03559; BAA04478.1;
CC PIR; T02340; T02340.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 7.
CC PRINTS; PR00320; GPROTEINRPT.
CC ProDom; PD0000018; WD40; 4.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00678; WD_REPEATS_1; 4.
CC PROSITE; PS00082; WD_REPEATS_2; 6.
CC PROSITE; PS0294; WD_REPEATS_REGION; 1.
CC REPEAT; WD repeat.
CC REPEAT 14 45 WD 1.
CC REPEAT 62 92 WD 2.
CC REPEAT 104 134 WD 3.
CC REPEAT 148 180 WD 4.
CC REPEAT 192 222 WD 5.
CC REPEAT 233 262 WD 6.
CC REPEAT 292 322 WD 7.
CC REPEAT 25 25 A -> P.
CC REPEAT 46 46 I -> L.
CC REPEAT 116 116 V -> A.
CC REPEAT 142 142 T -> I.
CC REPEAT 188 188 L -> A.
CC REPEAT 200 200 P -> T.
CC REPEAT 267 267 S -> T.
CC REPEAT 284 284 S -> A.
CC REPEAT 301 301 S -> G.
CC REPEAT 324 324 D -> G.
CC SEQUENCE 326 AA; 35945 MW; 680D8DA14AB6488D CRC64;
Query Match 37.9%; Score 39; DB 1; Length 326;
Best Local Similarity 42.9%; Pred. No. 54;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
OY 2 TTGTGVLKQDWG 15
Db ::|:|:|
73 SSDGMFALSGWDG 86
RESULT 34
YF02_MYCPN STANDARD; PRT; 422 AA.
AC P75285;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MPN502 (P02_orf422V).
GN MPN502 OR WP341.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,

```

```
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL Nucleic Acids Res. 24:4420-4449(1996).
RC -!- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000032; AAB95988.1; -
DR PIR; S73667; S73667.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 422 AA; 46141 MW; 03BE7D381C70C298 CRC64;

Query Match 37.9%; Score 39; DB 1; Length 422;
Best Local Similarity 70.0%; Pred. No. 71;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTTGVFGLK 10
   ||| | :|||
DB 141 NPTTIGAYGLK 150

RESULT 35
PYRC_BACCL
ID PYRC_BACCL STANDARD; PRT; 427 AA.
AC P46538;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydroorotase (EC 3.5.2.3) (DHOase).
GN PYRC.
OS Bacillus caldolyticus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1394;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DSM 405;
RC MEDLINE=94282293; PubMed=7516791;
RA Ghm S.Y., Nielsen P., Neuhaed J.;
RT "Molecular characterization of pyrimidine biosynthesis genes from the
RL thermophile Bacillus caldolyticus.";
RL Microbiology 140:479-491(1994).
CC -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -!- COFACTOR: This enzyme tightly binds one zinc atom per chain which
CC is required for the catalytic mechanism. It also binds weakly to
CC two other zincs which are not essential for activity (By
CC similarity).
CC -!- PATHWAY: Pyrimidine biosynthesis; third step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DHOASE FAMILY. SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X73308; CAA51737.1; -
DR PIR; I40167; I40167.
DR HANAP; MF_00220; -; 1.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR004722; PyrC_multi.
DR InterPro; IPR005847; Urease.
```

```
DR Pfam; PF01979; Amidohydro_1; 1.
DR ProDom; PD000518; Urease; 1.
DR TIGRFAMs; TIGR00857; PYRC_multi; 1.
DR PROSITE; PS00482; DIHYDROOROTASE_1; 1.
DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
KW Pyrimidine biosynthesis; Hydrolyase; Zinc.
FT METAL 60 62 ZINC (POTENTIAL).
FT METAL 60 62 ZINC (POTENTIAL).
SQ SEQUENCE 427 AA; 46047 MW; 759A2AA99F733F4E CRC64;

Query Match 37.9%; Score 39; DB 1; Length 427;
Best Local Similarity 75.0%; Pred. No. 72;
Matches 9; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

QY 4 TGVFGLKQ--DW 13
   |||| ||| ||
DB 342 TGVFGLKQLVDW 353

Search completed: October 6, 2003, 07:44:15
Job time : 6.38806 secs
```

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 10.209 Seconds
(without alignments)
178.981 Million cell updates/sec

Title: US-09-765-739A-7

Perfect score: 103

Sequence: 1 NTTFVFLKQDWDGSTRIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97	94.2	280	2 JE0217	28k surface antige
2	88	85.4	276	2 JE0218	28k surface antige
3	74	71.8	286	2 JE0219	28k surface antige
4	60	58.3	278	2 JE0216	28k surface antige
5	59	57.3	284	2 I40882	major antigenic pr
6	55	53.4	133	2 JE0221	28k surface antige
7	51	49.5	540	2 S54586	probable membrane
8	49	47.6	5188	2 B85547	probable RTX fami
9	49	47.6	5291	2 F90696	hypothetical prote
10	47	45.6	584	2 C48658	flagellin - Escher
11	46	44.7	1806	2 AF1717	probable peptidogl
12	44	42.7	534	2 C82096	aminoacyl-histidin
13	44	42.7	1649	2 C86822	hypothetical prote
14	43.5	42.2	267	2 C83242	conserved hypot
15	43.5	42.2	290	2 S76787	hypothetical prote
16	43	41.7	160	2 A75466	2-demethylmenaquin
17	43	41.7	185	2 S63457	hypothetical prote
18	43	41.7	490	2 G85354	hypothetical prote
19	43	41.7	648	1 P3BPF6	p3 protein - phage
20	42	40.8	160	2 A12476	hypothetical prote
21	42	40.8	307	2 AB2459	hypothetical prote
22	42	40.8	384	2 T40502	hypothetical prote
23	42	40.8	466	2 F95307	conserved hypot
24	42	40.8	634	2 F97172	flagellar hook-ass
25	42	40.8	1029	2 F87369	TonB-dependent rec
26	41.5	40.3	293	2 A84110	sugar ABC transpor
27	41.5	40.3	467	2 AE1162	hypothetical prote
28	41	39.8	118	2 S34346	hypothetical prote
29	41	39.8	130	2 C88102	protein W09G10.6 I

30	41	39.8	135	2 I56326	fatty acid binding
31	41	39.8	185	2 F75255	hypothetical prote
32	41	39.8	202	2 A95905	probable transcrip
33	41	39.8	244	2 E84885	hypothetical prote
34	41	39.8	335	2 B82220	hypothetical prote
35	41	39.8	375	2 AC1350	N-acetylmuramoyl-L
36	41	39.8	375	2 AF1720	N-acetylmuramoyl-L
37	41	39.8	478	2 AB0779	probable lipoprote
38	41	39.8	653	2 F85620	partial fibrillar u
39	41	39.8	653	2 H90756	partial fibrillar u
40	41	39.8	819	2 B87580	conserved hypot
41	41	39.8	866	1 C84834	probable outer mem
42	41	39.8	893	2 E95053	cell wall surface
43	40.5	39.3	622	2 AB1090	bacteriophage mino
44	40.5	39.3	1055	2 A96682	protein F1E22.12 I
45	40.5	39.3	1559	2 T07757	probable DNA (cyto
46	40	38.8	85	2 T27307	hypothetical prote
47	40	38.8	115	2 G86544	glycine cleavage s
48	40	38.8	115	2 C72079	glycine cleavage s
49	40	38.8	282	2 F64695	biotin synthetase
50	40	38.8	282	2 H71823	biotin synthetase
51	40	38.8	299	2 T23932	hypothetical prote
52	40	38.8	319	2 T36845	probable membrane
53	40	38.8	377	2 F69008	acetyltransferase
54	40	38.8	404	2 S77505	probable phosphos
55	40	38.8	408	2 S76830	hypothetical prote
56	40	38.8	470	2 JG4098	tetracycline 6-hyd
57	40	38.8	482	2 G75483	probable leucyl am
58	40	38.8	501	2 AG3356	glutamate-tRNA lig
59	40	38.8	585	2 F90961	flagellin importe
60	40	38.8	585	2 F85809	hypothetical prote
61	40	38.8	719	2 A83800	penicillin-binding
62	40	38.8	964	1 VCLJC6	env polypeptide pr
63	40	38.8	1025	2 A54718	dihydropyrimidine
64	40	38.8	1118	1 A49724	protein-tyrosine-p
65	40	38.8	1191	2 A53491	bumetanide-sensiti
66	40	38.8	4861	2 S71752	giant protein p619
67	39.5	38.3	212	2 C90905	hypothetical prote
68	39.5	38.3	216	2 B85712	unknown protein en
69	39.5	38.3	260	2 S74597	ABC-type transport
70	39.5	38.3	366	2 A81183	glycine cleavage s
71	39.5	38.3	368	2 G81919	probable aminometh
72	39.5	38.3	1545	2 T14288	DNA (cytosine-5')-
73	39	37.9	88	2 A84166	hypothetical prote
74	39	37.9	107	2 PH0999	ig heavy chain V r
75	39	37.9	149	2 C97180	uncharacterized pr
76	39	37.9	163	2 C81283	flavodoxin C1382c
77	39	37.9	189	2 G82586	translation elonga
78	39	37.9	231	2 T32146	hypothetical prote
79	39	37.9	233	2 S76857	hypothetical prote
80	39	37.9	245	2 F84680	hypothetical prote
81	39	37.9	258	2 E69019	conserved hypot
82	39	37.9	261	2 G84057	hypothetical prote
83	39	37.9	284	2 B69945	phage-related prot
84	39	37.9	290	2 T02300	GTP-binding regula
85	39	37.9	300	2 S60558	envelope polyprote
86	39	37.9	300	2 S60526	envelope polyprote
87	39	37.9	307	2 T44893	hypothetical prote
88	39	37.9	326	2 T16987	GTP-binding protei
89	39	37.9	326	2 T02340	GTP-binding protei
90	39	37.9	328	2 T16970	GTP-binding protei
91	39	37.9	383	2 G72359	hypothetical prote
92	39	37.9	390	2 AB1656	cystathionine beta
93	39	37.9	390	2 AG1284	cystathionine beta
94	39	37.9	422	2 S73667	adhesin P1 precurs
95	39	37.9	427	2 I40167	dihydroorotase (EC
96	39	37.9	428	2 S73379	adhesin P1 precurs
97	39	37.9	438	2 S73952	adhesin P1 precurs
98	39	37.9	459	2 AH0405	probable cytochrom
99	39	37.9	471	2 T46860	probable glutamate
100	39	37.9	473	2 JQ1936	xylan 1,4-beta-xy

ALIGNMENTS

RESULT 1

28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0217
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 94.2%; Score 97; DB 2; Length 280;
Best Local Similarity 94.7%; Pred. No. 8.6e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGDSGIS 19
||| |||||:||||| ||
Db 60 NTTGVFGLKQDWGDSGIS 78

RESULT 2

28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0218
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 85.4%; Score 88; DB 2; Length 276;
Best Local Similarity 84.2%; Pred. No. 2.3e-06;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGDSGIS 19
||| |||||:||||| ||
Db 59 NTTGVFGLKQDWGDSGIS 77

RESULT 3

28k surface antigen 2 - Ehrlichia chaifensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaifensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0219
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match 71.8%; Score 74; DB 2; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00039;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGDSGIS 19
||| |||||:||||| ||
Db 60 NTTGVFGLKQDWDRCVIS 78

RESULT 4

28k surface antigen 3 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0216
A:Molecule type: DNA
A:Residues: 1-278 <RED>
A:Cross-references: GB:AF062761

Query Match 58.3%; Score 60; DB 2; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.063;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGDSGIS 19
||| |||||:||||| ||
Db 60 NPTVALYGLKQDWEGISS 78

RESULT 5

I40882
Major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R:van Vliet, A.H.; Jongejan, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding th
A:Reference number: I40882; MUID:94178956; PMID:8132352
A:Accession: I40882
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:9454266; PIDN:CAA52309.1; PID:9454267
C:Genetics:
A:Gene: map1

Query Match 57.3%; Score 59; DB 2; Length 284;
Best Local Similarity 76.9%; Pred. No. 0.092;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTGVFGLKQDWG 15
| |||||:|||||
Db 63 TKAVFGLKKDWG 75

RESULT 6

JE0221
28k surface antigen 2 - Ehrlichia canis
C:Species: Ehrlichia canis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
C:Accession: JE0221
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; BurrIDGE, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0221
A:Molecule type: DNA
A:Residues: 1-133 <RED>
A:Cross-references: GB:AF062762; NID:93327964; PIDN:AAC26722.1; PID:93327966

Query Match 53.4%; Score 55; DB 2; Length 133;

R.White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioreistant bacterium *Deinococcus radiodurans* R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: A75466
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <WHI>
 A:Cross-references: GB:AE001940; GB:AE000513; NID:g6458577; PIDN:AAF10437.1; PID:g645857
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0859
 A:Map position: 1
 C:Keywords: methyltransferase

Query Match 41.7%; Score 43; DB 2; Length 160;
 Best Local Similarity 50.0%; Pred. No. 17;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGKQDWGSTI 18
 ||||| :|||
 DB 80 GVFGVNGWEGVII 93

RESULT 17
 S63457
 hypothetical protein YPL025c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein LPB6c
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 16-May-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002
 C:Accession: S63457
 R:Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.; V
 submitted to the EMBL Data Library, September 1995
 A:Reference number: S63452
 A:Accession: S63457
 A:Molecule type: DNA
 A:Residues: 1-185 <WAN>
 A:Cross-references: EMBL:U36624; NID:gl276642; PID:gl039452; GSPDB:GN00016; MIPS:YPL025c
 C:Genetics:
 A:Gene: MIPS:YPL025c
 A:Cross-references: SGD:S0005946
 A:Map position: 16L
 C:Superfamily: *Saccharomyces* hypothetical protein YPL025c

Query Match 41.7%; Score 43; DB 2; Length 185;
 Best Local Similarity 52.6%; Pred. No. 20;
 Matches 10; Conservative 1; Mismatches 2; Indels 6; Gaps 1;

QY 1 NPTTGVFLKQDWGSTIS 19
 ||| :|||
 DB 82 NQTRGM-----WDGSTIS 94

RESULT 18
 G85354
 hypothetical protein AT4g30340 [imported] - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
 C:Accession: G85354
 R:anonymus, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
 Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.
 A:Reference number: A85001; MUID:20083488; PMID:10617198
 A:Accession: G85354
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <STO>
 A:Cross-references: GB:NC_001268; NID:g7269934; PIDN:CAB81027.1; GSPDB:GN00140
 C:Genetics:
 A:Gene: AT4g30340
 A:Map position: 4

Query Match 41.7%; Score 43; DB 2; Length 490;
 Best Local Similarity 53.8%; Pred. No. 57;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 VFGLKQDWGSTI 18
 :||||| :
 DB 414 IFGLKQGWHSFV 426

RESULT 19
 P3BPF6
 p3 protein - phage phi-6
 C:Species: phage phi-6
 A:Note: host *Pseudomonas phaseolicola*
 C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 23-Jul-1999
 C:Accession: C28648
 R:Gottlieb, P.; Metzger, S.; Romantschuk, M.; Carton, J.; Strassman, J.; Bamford, D.H
 Virology 163, 183-190, 1988
 A:Title: Nucleotide sequence of the middle dsRNA segment of bacteriophage phi-6: plac
 A:Reference number: A94372; MUID:88160044; PMID:3347997
 A:Accession: C28648
 A:Molecule type: genomic RNA
 A:Residues: 1-648 <GOT>
 A:Cross-references: GB:M17462; NID:g862621; PIDN:AAA68485.1; PID:g215490
 C:Comment: The genome of this phage consists of three segments of double-stranded RNA
 C:Comment: This protein is required for adsorption onto host cells.
 C:Genetics:
 A:Gene: 3
 A:Map position: segment M
 C:Superfamily: phage phi-6 p3 protein
 C:Keywords: late protein

Query Match 41.7%; Score 43; DB 1; Length 648;
 Best Local Similarity 43.8%; Pred. No. 78;
 Matches 7; Conservative 5; Mismatches 2; Indels 2; Gaps 1;

QY 6 VFG--LKQDWGSTIS 19
 :|| :||| :
 DB 181 IFGWYVKWDEGSAVA 196

RESULT 20
 A12476
 hypothetical protein all5369 [imported] - *Nostoc* sp. (strain PCC 7120)
 C:Species: *Nostoc* sp. PCC 7120
 A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: A12476
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irig
 Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
 A:Reference number: ABL807; MUID:21595285; PMID:11759840
 A:Accession: A12476
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-160 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BAF77068.1; PID:gl17134508; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: all5369

Query Match 40.8%; Score 42; DB 2; Length 160;
 Best Local Similarity 77.8%; Pred. No. 25;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 GVFGKQDW 13
 ||| |||||
 DB 117 GVFSLKQNW 125

RESULT 21

AB2459
 hypothetical protein alr5226 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AB2459
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AB2459
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BA076925.1; PID:gl7134365; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr5226

Query Match 40.8%; Score 42; DB 2; Length 307;
 Best Local Similarity 47.1%; Pred. No. 50;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 TTGTGVLKQDWGSTI 18
 I : | | | | |
 Db 240 TMSHALGAKPKWGSTI 256

RESULT 22

T40502
 hypothetical protein SPBC4F6.05c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T40502
 R:Gallagher, R.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21933
 A:Accession: T40502
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-584 <GWI>
 A:Cross-references: EMBL:AL031534; PIDN:CAA20725.1; GSPDB:GN00067; SPDB:SPBC4F6.05c
 A:Experimental source: strain 972h-; cosmid c4F6
 C:Genetics:
 A:Gene: SPDB:SPBC4F6.05c
 A:Map position: 2

Query Match 40.8%; Score 42; DB 2; Length 384;
 Best Local Similarity 57.1%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 VFGKQDWGSTIS 19
 I | | | | |
 Db 110 VFGASDKWGLLIS 123

RESULT 23

F95307
 conserved hypothetical protein Sma0690 [imported] - Sinorhizobium meliloti (strain 1021)
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: F95307
 R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
 ; Kalman, S.; Keating, D.H.; Palm, C.; Beck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
 A:Reference number: A95262; MUID:21396503; PMID:11481432
 A:Accession: F95307
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-466 <KUR>
 A:Cross-references: GB:AE006469; PIDN:AAK65024.1; PID:gl4523454; GSPDB:GN00165

A:Experimental source: strain 1021, megaplasmid pSymA
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler, F.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
 rebault, P.; Vandenbol, M.; Vorholter, R.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Sma0690
 A:Genome: plasmid

Query Match 40.8%; Score 42; DB 2; Length 466;
 Best Local Similarity 70.0%; Pred. No. 76;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 10 KQDWGSTIS 19
 I : | | | | |
 Db 96 KESWDGVSIS 105

RESULT 24

F97172
 flagellar hook-associated protein FlgK [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: F97172
 R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: F97172
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-634 <KUR>
 A:Cross-references: GB:AF001437; PIDN:AAK80169.1; PID:gl5025209; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC2212

Query Match 40.8%; Score 42; DB 2; Length 634;
 Best Local Similarity 50.0%; Pred. No. 11e+02;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 NTTGVLKQDWGSTI 18
 I | | | | | | | | | |
 Db 536 DTDGVTIKSDPDGVTV 553

RESULT 25

F87369
 TonB-dependent receptor [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: F87369
 R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Neilson, K.E.; Eisen, J.; Heidelberg,
 N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: F87369; MUID:21173698; PMID:11259647
 A:Accession: F87369
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1029 <STO>
 A:Cross-references: GB:AE005673; NID:gl3422250; PIDN:AAK22954.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC0970

Query Match 40.8%; Score 42; DB 2; Length 1029;

Best Local Similarity 50.0%; Pred. NO. 1.9e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY 1 NTTTGVFGL--KQDWGGS 16
| : : || || || || ||
Db 219 NNSGCVGLSGNNDWGG 236

RESULT 26
A84110
sugar ABC transporter (permease) BH3681 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A84110
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A84110
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-293 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:gl0176109; PIDN:BA07400.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3681
C:Superfamily: inner membrane protein ugpA

Query Match 40.3%; Score 41.5; DB 2; Length 293;
Best Local Similarity 66.7%; Pred. NO. 57;
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 4 TGVFGLKQDWG 15
|| : || || || ||
Db 33 TGYVGL-MDWG 43

RESULT 27
AE1162
hypothetical protein lmo0701 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AE1162
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1162
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-467 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC98779.1; PID:gl16410090; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0701

Query Match 40.3%; Score 41.5; DB 2; Length 467;
Best Local Similarity 52.6%; Pred. NO. 94;
Matches 10; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 2 TTTGVFGLK-QDWGSGTIS 19
||||| | ||||| |
Db 361 TTTGKITLHWTDWDGQVFS 379

RESULT 28
S34346
hypothetical protein B - Shigella flexneri (fragment)
C:Species: Shigella flexneri

C:Date: 22-Nov-1993 #sequence_revision 12-May-1995 #text_change 12-May-1995
C:Accession: S34346
R:Faubladiet, M.; Bouche, J.P.
submitted to the EMBL Data Library, June 1993
A:Description: Genetic elements related to bacteriophages Lambda and P4 among a wides
A:Reference number: S34326
A:Accession: S34346
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-118 <FAU>
A:Cross-references: EMBL:Z23100

Query Match 39.8%; Score 41; DB 2; Length 118;
Best Local Similarity 46.7%; Pred. NO. 25;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 5 GVFGKQDWGSGTIS 19
| | | ||| || : :
Db 55 GCAGSQDWPGRV 69

RESULT 29
C88102
protein W09G10.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C88102
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: C88102
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-130 <STO>
A:Cross-references: GB:chr_II; PIDN:AB66114.1; PID:g2315670; GSPDB:GN00020; CESP:W09G
C:Genetics:
A:Gene: W09G10.6
A:Map position: 2

Query Match 39.8%; Score 41; DB 2; Length 130;
Best Local Similarity 57.1%; Pred. NO. 28;
Matches 8; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

QY 2 TTTGVFGLKQDWG 15
: ||| : | : |||
Db 83 STTGIDGI--DWDG 94

RESULT 30
I56326
fatty acid binding protein homolog - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I56326
R:Madsen, P.; Rasmussen, H.H.; Leffers, H.; Honore, B.; Celis, J.E.
J. Invest. Dermatol. 99, 299-305, 1992
A:Title: Molecular cloning and expression of a novel keratinocyte protein (psoriasis-
ilarity to fatty acid-binding proteins.
A:Reference number: I56326; MUID:92381332; PMID:1512466
A:Accession: I56326
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-135 <RES>
A:Cross-references: GB:M94856; NID:g182353; PIDN:AAA58467.1; PID:g182354
C:Genetics:
A:Gene: PA-FABP
C:Superfamily: myelin P2 protein

Query Match 39.8%; Score 41; DB 2; Length 135;
Best Local Similarity 47.6%; Pred. NO. 29;

Matches 10; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 NTGTGVLKQDWG--STIS 19
| | | | | | | | | |
Db 88 NFTDGLVQHQEWGDKRESTIT 108

RESULT 31

hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
C:Accession: F75255
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-185 <WHI>
A:Cross-references: GB:AE002089; GB:AE000513; NID:g6460427; PIDN:AAF12137.1; PID:g646042
A:Experimental source: strain R1
C:Genetics:
A:Map position: 1
A:Gene: DR2593
C:Superfamily: Deinococcus radiodurans hypothetical protein DR1748

Query Match 39.8%; Score 41; DB 2; Length 185;
Best Local Similarity 43.8%; Pred. No. 41;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 TGVFLKQDWGSGTIS 19
| | | | | | | | | |
Db 31 TAGYDLHADWGSRLT 46

RESULT 32

probable transcription regulator protein [Imported] - Sinorhizobium meliloti (strain 102)
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: A95905
R:Finan, T.W.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A95905
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC48905.1; PID:GI5140378; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Bariloy-Hubler,
pela, D.; Chai, F.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contact: annotation
C:Genetics:
A:Gene: Smb20525
A:Genome: plasmid

Query Match 39.8%; Score 41; DB 2; Length 202;
Best Local Similarity 50.0%; Pred. No. 46;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 4 TGVFLKQDWGSGTIS 19
| | | | | | | | | |

Db 122 TEIFGLYRWLAKOIS 137
| : | | | : | | | | |

RESULT 33

hypothetical protein Atg45010 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84885
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
euss, D.; Nierman, W.C.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84885
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <STO>
A:Cross-references: GB:AE002093; NID:g4895249; PIDN:AAD32834.1; GSPDB:GN00139
A:Genetics:
A:Gene: AT2g45010
A:Map position: 2

Query Match 39.8%; Score 41; DB 2; Length 244;
Best Local Similarity 58.3%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTGTGVLKQD 12
| | | | | | | | | |
Db 61 NWTGTGFGCAED 72

RESULT 34

hypothetical protein VC1266 [Imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Basu, S.; Qin, H.; Dragoi, I.; Sellers
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: B82220
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <HEI>
A:Cross-references: GB:AE004206; GB:AE003852; NID:g9655749; PIDN:AAF94425.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1266
A:Map position: 1

Query Match 39.8%; Score 41; DB 2; Length 335;
Best Local Similarity 50.0%; Pred. No. 79;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 TTGVLKQDWGDS 16
| | | | | | | | | |
Db 59 TTDISALKQEWQAS 72

RESULT 35

N-acetylmuramoyl-L-alanine amidase and to internalin B homolog lmo2203 [Imported] - L
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC1350
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihi,

D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AC1350
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-375 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAD00281.1; PID:g16411673; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo2203

Query Match 39.8%; Score 41; DB 2; Length 375;
 Best Local Similarity 33.3%; Pred No. 89;
 Matches 6; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
 QY 2 TTTGVFLKQDWDGSTIT 19
 : :||:| :||:|
 Db 95 SANNLEGIKGSYEGSSVS 112

Search completed: October 6, 2003, 07:49:44
 Job time : 12.209 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:03 ; Search time 31.8358 Seconds
(without alignments)
89.744 Million cell updates/sec

Title: US-09-765-739A-6

Perfect score: 97

Sequence: 1 NPTVALYGLKQDWEGISS 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_19Jun03.*

1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	97	100.0	18	ABG30748
2	97	100.0	278	19
3	97	100.0	278	20
4	97	100.0	278	21
5	97	100.0	278	22
6	97	100.0	278	23
7	97	100.0	278	24
8	97	100.0	278	25
9	97	100.0	278	26

10	88	90.7	18	23	ABG30746	Ehrlichia chaffeen
11	88	90.7	280	20	AAU06945	E. chafeensis OMP-
12	88	90.7	280	23	ABG77937	Ehrlichia chaffeen
13	88	90.7	280	23	AAU96107	Ehrlichia chafeens
14	88	90.7	280	23	AAU73414	Ehrlichia chaffeen
15	62	63.9	280	19	AAW51094	Ehrlichia chaffeen
16	62	63.9	280	21	AAAB36188	Ehrlichia chaffeen
17	62	63.9	280	22	AAU04198	Variable surface a
18	62	63.9	280	23	AAU73417	Ehrlichia chaffeen
19	59	60.8	19	23	ABG30749	Ehrlichia chaffeen
20	59	60.8	20	23	ABG30744	Ehrlichia canis pe
21	59	60.8	280	20	AAU06948	E. chafeensis OMP-
22	59	60.8	280	23	ABG77940	Ehrlichia chaffeen
23	59	60.8	280	23	AAU96110	Ehrlichia chafeens
24	59	60.8	288	20	AAU06959	E. canis P30 prote
25	59	60.8	288	23	ABG77950	Ehrlichia canis ou
26	58	59.8	20	23	ABG30745	Ehrlichia chaffeen
27	58	59.8	256	20	AAU06942	E. chafeensis p28
28	58	59.8	256	23	ABG77966	Protein encoded by
29	58	59.8	276	19	AAW51095	Ehrlichia chaffeen
30	58	59.8	276	21	AAAB36189	Ehrlichia chaffeen
31	58	59.8	276	22	AAU04199	Variable surface a
32	58	59.8	280	19	AAW51089	Ehrlichia chaffeen
33	58	59.8	280	21	AAAB36183	Ehrlichia chaffeen
34	58	59.8	280	22	AAU04193	Major antigenic pr
35	58	59.8	281	20	AAU06943	E. chafeensis OMP-
36	58	59.8	281	23	ABG77935	Ehrlichia chaffeen
37	58	59.8	281	23	AAU96105	Ehrlichia chafeens
38	58	59.8	280	23	AAU73418	Ehrlichia chaffeen
39	56	57.7	280	23	AAU06962	E. canis P30-2 pro
40	56	57.7	280	21	AAU71479	Ehrlichia canis im
41	56	57.7	280	23	ABG77953	Ehrlichia canis ou
42	56	57.7	280	23	AAU96102	Ehrlichia canis p2
43	56	57.7	284	23	AAU96111	Cowdria ruminantii
44	53	54.6	20	23	ABG30743	Ehrlichia canis pe
45	53	54.6	278	21	AAU71477	Ehrlichia canis im
46	53	54.6	278	23	AAU96100	Ehrlichia canis p2
47	53	54.6	307	20	AAU06961	E. canis P30-1 pro
48	53	54.6	307	23	ABG77952	Ehrlichia canis ou
49	52	53.6	287	19	AAW51088	Cowdria ruminantii
50	52	53.6	287	21	AAAB36182	Cowdria ruminantii
51	52	53.6	287	22	AAU04192	Major antigenic pr
52	51	52.6	19	23	ABG30747	Ehrlichia chaffeen
53	51	52.6	286	19	AAW51092	Ehrlichia chaffeen
54	51	52.6	286	20	AAU06946	E. chafeensis OMP-
55	51	52.6	286	21	AAAB36186	Ehrlichia chaffeen
56	51	52.6	286	22	AAU04196	Variable surface a
57	51	52.6	286	23	ABG77938	Ehrlichia chaffeen
58	51	52.6	286	23	AAU96108	Ehrlichia chafeens
59	51	52.6	286	23	AAU73415	Ehrlichia chaffeen
60	48	49.5	132	22	AAU04201	Variable surface a
61	48	49.5	133	19	AAW51097	Ehrlichia canis VS
62	48	49.5	133	21	AAAB36191	Ehrlichia canis pa
63	48	49.5	133	21	AAU71480	Ehrlichia canis im
64	48	49.5	133	23	AAU96103	Ehrlichia canis p2
65	48	49.5	276	20	AAU06964	E. canis P30-4 pro
66	48	49.5	276	23	ABG77955	Ehrlichia canis ou
67	48	49.5	276	23	AAU96117	Ehrlichia canis p2
68	48	49.5	283	21	AAU71478	Ehrlichia canis im
69	48	49.5	283	23	AAU96101	Ehrlichia canis p2
70	47	48.5	65	22	AAU63107	Propionibacterium
71	47	48.5	308	22	AAH58819	Drosophila melanog
72	47	48.5	904	20	AAU21976	Senescence-associ
73	45	46.4	1242	22	ABB63759	Drosophila melanog
74	44	45.4	64	22	ABU15750	Human nervous syst
75	43	44.3	50	21	AAAB33357	Pinus radiata tran
76	43	44.3	53	21	AAAB33358	Pinus radiata tran
77	43	44.3	114	23	ABB79299	Human ovary specif
78	43	44.3	210	22	ABG22152	Novel human diagno
79	43	44.3	237	21	AAAB33257	Pinus radiata tran
80	43	44.3	759	22	ABG20610	Novel human diagno
81	42	43.3	180	23	ABG64967	Human albumin fusi
82	42	43.3	180	23	ABG64969	Human albumin fusi

83 Human gene 6 encod
84 Human gene 6 encod
85 Streptococcus pneu
86 S. pneumoniae type
87 Novel human diagno
88 Bacillus deramific
89 Bacillus deramific
90 Bacillus deramific
91 Bacillus deramific
92 Bacillus deramific
93 Bacillus deramific
94 Bacillus deramific
95 Bacillus deramific
96 Bacillus deramific
97 Bacillus deramific
98 Bacillus deramific
99 Bacillus deramific
100 Bacillus deramific

SQ Sequence 18 AA;
Query Match 100.0%; Score 97; DB 23; Length 18;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
| | | | | | | | | | | | | | | | | |
DB 1 NPTVALYGLKQDWEGISS 18

RESULT 2

AAW51093
ID AAW51093 standard; Protein; 278 AA.

XX AAW51093;

DT 14-SEP-1998 (first entry)

XX Ehrlichia chaffeensis VSA3 protein.

XX MAP1 homologue; variable surface antigen; VSA3; rickettsia;

KW DNA vaccine.

XX Ehrlichia chaffeensis.

XX Key Location/Qualifiers

FT Peptide 1..25

XX /note= "putative signal peptide"

PN WO9816554-A1.

XX 23-APR-1998.

PF 17-OCT-1997; 97WO-US19044.

PR 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

PI Barbet AF, Burrldge MJ, Ganta RR, Mahan SM, McGuire TC;

XX Nyika A, Rurangirwa FR;

DR WPI: 1998-251232/22.

DR N-PSDB; AAV07179.

PT Composition containing nucleic acid encoding rickettsial antigen -

PT useful for, e.g. stimulating protective immune response in humans or

PT animals

XX Claim 3; Fig 2A-B; 39pp; English.

CC This is the full-length variable surface antigen VSA3 protein of
CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
CC partial open reading frame (ORF5) of a genomic locus (see AAV07179)
CC of E. chaffeensis that was obtained on the basis of homology to the
CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
CC This genomic locus included 5 ORFs encoding similar, but
CC non-identical proteins (see AAW51091-95). A claimed composition
CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
CC (see AAW51088-99) that elicits a protective immune response against a
CC rickettsial pathogen. The nucleic acid is used, in human or
CC veterinary medicine, in vaccines to protect against Rickettsia,
CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
CC polypeptides can be used diagnostically to detect antibodies
CC associated with Ehrlichia infection (claimed).

XX Sequence 278 AA;

Query Match 100.0%; Score 97; DB 19; Length 278;
Best Local Similarity 100.0%; Pred. No. 4.7e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ALIGNMENTS

RESULT 1

ABG30748

ID ABG30748 standard; Peptide; 18 AA.

XX ABG30748;

DT 21-OCT-2002 (first entry)

XX Ehrlichia chaffeensis peptide fragment #4.

XX Antibody detection; monoclonal antibody; polyclonal antibody.

XX Ehrlichia chaffeensis.

PN WO200257794-A2.

XX 25-JUL-2002.

PF 16-JAN-2002; 2002WO-US01395.

PR 18-JAN-2001; 2001US-0765739.

XX (IDEXX-) IDEXX LAB INC.

PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;

DR WPI: 2002-599730/64.

PT New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals

PS Claim 1; Page 5; 29pp; English.

CC The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. chaffeensis peptide
CC fragment used in the composition of the invention.

QY 1 NPTVALYGLKQDWEGISS 18
 DB 60 NPTVALYGLKQDWEGISS 77

RESULT 3

AA06947
 ID AAY06947 standard; Protein; 278 AA.

AC AAY06947;

DT 05-JUL-1999 (first entry)

DE E. chaffeensis OMP-1E protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

OS Ehrlichia chaffeensis.

PN WO9913720-Al.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

DR WPI: 1999-254290/21.

DR N-PSDB; AAX34747.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis

PS Claim 15; Fig 7B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 278 AA;

Query Match 100.0%; Score 97; DB 20; Length 278;

Best Local Similarity 100.0%; Pred. No. 4.7e-08;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
 DB 60 NPTVALYGLKQDWEGISS 77

RESULT 4

AAB36187
 ID AAB36187 standard; Protein; 278 AA.

AC AAB36187;

DT 02-WAR-2001 (first entry)

DE Ehrlichia chaffeensis partial VSA3.

XX Ehrlichia chaffeensis; VSA3; variable surface antigen 3; MAP1;

KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;

KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;

KW 3gdorf3.

XX

OS Ehrlichia chaffeensis.

PN WO200065063-A2.

PD 02-NOV-2000.

PF 21-APR-2000; 2000WO-US10886.

PR 22-APR-1999; 99US-0130725.

PA (UYFL) UNIV FLORIDA.

PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;

DR WPI: 2000-679675/66.

DR N-PSDB; AAC68704.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

PS Claim 3; Page 44-45; 63pp; English.

CC The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.

SQ Sequence 278 AA;

Query Match 100.0%; Score 97; DB 21; Length 278;

Best Local Similarity 100.0%; Pred. No. 4.7e-08;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
 DB 60 NPTVALYGLKQDWEGISS 77

RESULT 5

AAU04197

ID AAU04197 standard; Protein; 278 AA.

AC AAU04197;

DT 23-OCT-2001 (first entry)

DE Variable surface antigen 3 (VSA3) from Ehrlichia chaffeensis.

KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.

OS Ehrlichia chaffeensis.

PN US6251872-B1.

PD 26-JUN-2001.

PF 17-OCT-1997; 97US-0953326.

PR 17-OCT-1996; 96US-0733230.

PA (UYFL) UNIV FLORIDA.

```

XX PI Barbet AF, Ganta RR, McGuire TC, Burrledge MJ, Nyika A;
XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX XX WPI; 2001-424487/45.
XX DR N-PSDB; AAS07578.
XX XX New MAP2 genes and polypeptides useful as vaccines for conferring
XX PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX PT as molecular markers in nucleic acid analysis procedures -
XX PS Example 3; Fig 2A-2B; 30pp; English.
XX CC The sequence represents the amino acid sequence of variable surface
XX CC antigen 3 (VSA3) isolated from Ehrlichia chaffeensis which
XX CC has similarity to major antigen protein (MAP). The MAP polynucleotides
XX CC and polypeptides are useful as vaccines for conferring immunity to
XX CC rickettsia infection, including Cowdria ruminantium causing heartwater.
XX CC The MAP polynucleotides may be used as molecular markers in nucleic acid
XX CC analysis procedures, and to produce the MAP polypeptides, which may
XX CC be used to raise antibodies that are reactive with the polypeptides.
XX CC The nucleic acids may further be used as probes to identify
XX CC complementary sequences within other nucleic acid molecules or genomes,
XX CC where such probes can be applied to identify or distinguish infectious
XX CC strains of organisms in diagnostic procedures or in rickettsial
XX CC research where identification of particular organisms or strains is
XX CC needed.
XX XX Sequence 278 AA;
XX XX
XX XX Query Match 100.0%; Score 97; DB 22; Length 278;
XX XX Best Local Similarity 100.0%; Pred. No. 4.7e-08;
XX XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX
XX QY 1 NPTVALYGLKQDWEGISS 18
XX DB |||||
XX DB 60 NPTVALYGLKQDWEGISS 77
XX
XX RESULT 6
XX ID AAE31090 standard; Protein; 278 AA.
XX XX
XX AC AAE31090;
XX XX
XX DT 24-FEB-2003 (first entry)
XX XX
XX DE Ehrlichia ruminantium protein of the invention.
XX XX
XX KW Vaccine; immunity; rickettsial infection; spotted fever; heart water;
XX KW typhus; pathogen; immunostimulant; antibacterial.
XX XX
XX OS Ehrlichia ruminantium.
XX XX
XX PN WO200266652-A2.
XX XX
XX PD 29-AUG-2002.
XX XX
XX PF 20-FEB-2002; 2002WO-US05772.
XX XX
XX PR 20-FEB-2001; 2001US-269944P.
XX XX
XX PA (UYFL ) UNIV FLORIDA.
XX XX
XX PI Barbet AF, Whitmore WW, Kamper SM, Simbi BH, Ganta RR;
XX PI Moreland AL, Mwangi DM, McGuire TC, Mahan SM;
XX XX WPI; 2002-723186/78.
XX XX
XX XX New Ehrlichia ruminantium polynucleotides, useful as vaccines for
XX PT inducing protective immunity, and protecting animals or humans against
XX PT rickettsial diseases, e.g. typhus, spotted fever or heart water -
XX XX
PS PS Disclosure; Page 56-57; 206pp; English.
PS XX The present invention relates to nucleic acid vaccines for conferring
PS CC immunity to rickettsial infection, including Ehrlichia ruminantium
PS CC (formerly Cowdria ruminantium). The invention also relates to novel
PS CC E. ruminantium polynucleotides and their corresponding proteins.
PS CC Sequences of the invention are useful for inducing immunity, particularly
PS CC protective immunity. They are also useful for detecting the presence of
PS CC E. ruminantium in a biological sample. They are useful in vaccines for
PS CC protecting animals or humans against rickettsial diseases, e.g. typhus,
PS CC spotted fever or heart water. Sequences of the invention are useful for
PS CC detecting antibodies to pathogens. The present sequence is E. ruminantium
PS CC protein of the invention.
PS XX
PS XX Sequence 278 AA;
PS XX
PS XX Query Match 100.0%; Score 97; DB 23; Length 278;
PS XX Best Local Similarity 100.0%; Pred. No. 4.7e-08;
PS XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
PS XX
PS QY 1 NPTVALYGLKQDWEGISS 18
PS DB |||||
PS DB 60 NPTVALYGLKQDWEGISS 77
PS
PS RESULT 7
PS ID ABG77939 standard; Protein; 278 AA.
PS XX
PS AC ABG77939;
PS XX
PS DT 15-NOV-2002 (first entry)
PS XX
PS DE Ehrlichia chaffeensis outer membrane protein (OMP) #5.
PS XX
PS KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
PS XX
PS OS Ehrlichia chaffeensis.
PS XX
PS PN US2002120115-A1.
PS XX
PS PD 29-AUG-2002.
PS XX
PS PF 28-JAN-2002; 2002US-0059964.
PS XX
PS PR 19-MAY-1999; 99US-0314701.
PS XX
PS PA (RIKI/) RIKIHISA Y.
PS PA (OHAS/) OHASHI N.
PS XX
PS PI Rikihisa Y, Ohashi N;
PS XX
PS DR WPI; 2002-618954/66.
PS DR N-PSDB; ABS63280.
PS XX
PS PT Isolated polynucleotide encoding an outer membrane protein of E. canis
PS PT or E. chaffeensis used in the diagnosis of infection -
PS PS
PS PS Disclosure; Fig 7B; 49pp; English.
PS XX
PS XX The invention relates to an isolated polynucleotide encoding an outer
PS CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
PS CC in the diagnosis of infection. An infection such as human ehrlichiosis or
PS CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
PS CC patient, providing a polypeptide or mixture of polypeptides, contacting
PS CC the sample with the polypeptide and assaying for the formation of a
PS CC complex between antibodies in the serum sample and the polypeptide, where
PS CC formation of a complex is indicative of infection with E. chaffeensis.
PS CC This sequence represents an Ehrlichia outer membrane protein of the
PS CC invention.
PS XX
PS XX Sequence 278 AA;

```

Query Match 100.0%; Score 97; DB 23; Length 278;
Best Local Similarity 100.0%; Pred. No. 4.7e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
Db | | | | | | | | | | | | | | | | | |
60 NPTVALYGLKQDWEGISS 77

RESULT 8
AAU96109
ID AAU96109 standard; Protein; 278 AA.

XX AC AAU96109;

XX 02-JUL-2002 (first entry)

XX Ehrlichia chaffeensis OMP-1E.

XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;

KW antibacterial.

XX Ehrlichia chaffeensis.

OS WO20022782-A2.

PN 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28759.

XX 12-SEP-2000; 2000US-0660587.

XX (RERE-) RES DEV FOUND.

XX Walker DH, Yu X, McBride JW;

XX WPI; 2002-351882/38.

XX New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -

PS Example 3; Figure 3; 106pp; English.

XX The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.

XX Sequence 278 AA;

Query Match 100.0%; Score 97; DB 23; Length 278;
Best Local Similarity 100.0%; Pred. No. 4.7e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
Db | | | | | | | | | | | | | | | | | |
60 NPTVALYGLKQDWEGISS 77

RESULT 9
AAU73416
ID AAU73416 standard; Protein; 278 AA.

XX AC AAU73416;

XX 12-MAR-2002 (first entry)

XX Ehrlichia chaffeensis outer membrane protein p28-17.

XX Ehrlichia; outer membrane protein; p28; antibiotic; vaccine.

XX Ehrlichia chaffeensis.
OS WO200183699-A2.

PN 08-NOV-2001.

XX 01-MAY-2001; 2001WO-US13997.

XX 01-MAY-2000; 2000US-201035P.

XX (RERE-) RES DEV FOUND.

XX Walker DH, Yu X;

XX WPI; 2002-066527/09.

XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT p28 useful as a vaccine against Ehrlichia chaffeensis -

XX Disclosure; Figure 2; 97pp; English.

XX The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (p28-1 to p28-21) of Ehrlichia chaffeensis. p28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. p28
CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding p28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis p28 outer membrane proteins of the invention.

XX Sequence 278 AA;

Query Match 100.0%; Score 97; DB 23; Length 278;

Best Local Similarity 100.0%; Pred. No. 4.7e-08;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18

Db | | | | | | | | | | | | | | | | | |
60 NPTVALYGLKQDWEGISS 77

RESULT 10

ABG30746

ID ABG30746 standard; Peptide; 18 AA.

XX AC ABG30746;

XX 21-OCT-2002 (first entry)

XX Ehrlichia chaffeensis peptide fragment #2.

XX Antibody detection; monoclonal antibody; polyclonal antibody.

XX Ehrlichia chaffeensis.

XX WO200257794-A2.

XX 25-JUL-2002.

XX 16-JAN-2002; 2002WO-US01395.

XX 18-JAN-2001; 2001US-0765739.

XX (IDEX-) IDEXX LAB INC.

XX Lawton R, O'Connor TP, Bartol BA, Machenry PS;

XX WPI; 2002-599730/64.

XX New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals

XX

PS Claim 1; Page 5; 29pp; English.

CC The invention relates to a composition of matter comprising a polypeptide
 CC isolated from Ehrlichia species. The composition can be used for
 CC detecting the presence of antibodies to Ehrlichia, comprising contacting
 CC one or more polypeptides with a test sample suspected of comprising
 CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 CC complexes to form and detecting the complexes, where the detection of
 CC polypeptide/antibody complexes is an indication that antibodies to
 CC Ehrlichia are present in the test sample. The composition is useful for
 CC detecting or quantifying the presence of E. canis or E. chaffeensis
 CC infection in mammals. The polypeptides can be used to develop monoclonal
 CC and/or polyclonal antibodies that can be employed in assay systems and in
 CC the generation of chimeric antibodies for therapeutic use or other
 CC similar applications. This sequence represents an E. chaffeensis peptide
 CC fragment used in the composition of the invention.

XX Sequence 18 AA;

Query Match 90.7%; Score 88; DB 23; Length 18;
 Best Local Similarity 83.3%; Pred. NO. 7e-08; Mismatches 1; Indels 0; Gaps 0;
 Matches 15; Conservative 2;

QY 1 NPTVALYGLKQDWEGISS 18
 |||||
 Db 1 NPTVALYGLKQDWNGVSA 18

RESULT 11

AY06945
 ID AAY06945 standard; Protein; 280 AA.

AC AAY06945;

DT 05-JUL-1999 (first entry)

DE E. chaffeensis OMP-1C protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

OS Ehrlichia chaffeensis.

PN WO9913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

DR WPI; 1999-254290/21.

DR N-PSDB; AAX34745.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and

PT Ehrlichia canis

PS Claim 13; Fig 5B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 280 AA;

Query Match 90.7%; Score 88; DB 20; Length 280;

Best Local Similarity 83.3%; Pred. NO. 1.6e-06;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
 |||||
 Db 60 NPTVALYGLKQDWNGVSA 77

RESULT 12

ABG77937
 ID ABG77937 standard; Protein; 280 AA.

XX ABG77937;

DT 15-NOV-2002 (first entry)

DE Ehrlichia chaffeensis outer membrane protein (OMP) #3.

KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

OS Ehrlichia chaffeensis.

PN US2002120115-A1.

PD 29-AUG-2002.

PF 28-JAN-2002; 2002US-0059964.

PR 19-MAY-1999; 99US-0314701.

PA (RIKI/) RIKIHISA Y.

PA (OHAS/) OHASHI N.

PI Rikihisa Y, Ohashi N;

DR WPI; 2002-618954/66.

DR N-PSDB; ABS63278.

PT Isolated polynucleotide encoding an outer membrane protein of E. canis

PT or E. chaffeensis used in the diagnosis of infection -

PS Disclosure; Fig 5B; 49pp; English.

CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.

XX Sequence 280 AA;

Query Match 90.7%; Score 88; DB 23; Length 280;
 Best Local Similarity 83.3%; Pred. NO. 1.6e-06;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18
 |||||
 Db 60 NPTVALYGLKQDWNGVSA 77

RESULT 13

AAU96107
 ID AAU96107 standard; Protein; 280 AA.

XX AAU96107;

DT 02-JUL-2002 (first entry)

DE Ehrlichia chaffeensis OMP-1C.

XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
KW antibacterial.

XX Ehrlichia chaffeensis.

XX WO200222782-A2.

XX 21-MAR-2002.

XX PF 12-SEP-2001; 2001WO-US28759.

XX PR 12-SEP-2000; 2000US-0660587.

XX PA (RERE-) RES DEV FOUND.

XX PI Walker DH, Yu X, McBride JW;

XX DR WPI; 2002-351882/38.

XX New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -

XX Example 3; Figure 3; 106pp; English.

XX The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.

XX Sequence 280 AA;

Query Match 90.7%; Score 88; DB 23; Length 280;

Best Local Similarity 83.3%; Pred. No. 1.6e-06;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18

DB 60 NPTVALYGLKQDWNGVSA 77

RESULT 14

AAU73414

ID AAU73414 standard; Protein; 280 AA.

XX AC AAU73414;

XX DT 12-MAR-2002 (first entry)

XX DE Ehrlichia chaffeensis outer membrane protein p28-15.

XX KW Ehrlichia; outer membrane protein; p28; antibiotic; vaccine.

XX OS Ehrlichia chaffeensis.

XX PN WO200183699-A2.

XX PD 08-NOV-2001.

XX PF 01-MAY-2001; 2001WO-US13997.

XX PR 01-MAY-2000; 2000US-201035P.

XX PA (RERE-) RES DEV FOUND.

XX PI Walker DH, Yu X;

XX DR WPI; 2002-066527/09.

XX

PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT p28 useful as a vaccine against Ehrlichia chaffeensis -

XX Disclosure; Figure 2; 97pp; English.

XX The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (p28-1 to p28-21) of Ehrlichia chaffeensis. p28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. p28
CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding p28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis p28 outer membrane proteins of the invention.

XX Sequence 280 AA;

Query Match 90.7%; Score 88; DB 23; Length 280;

Best Local Similarity 83.3%; Pred. No. 1.6e-06;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEGISS 18

DB 60 NPTVALYGLKQDWNGVSA 77

RESULT 15

AAW51094

ID AAW51094 standard; Protein; 280 AA.

XX AC AAW51094;

XX DT 14-SEP-1998 (first entry)

XX DE Ehrlichia chaffeensis VSA4 protein.

XX KW MAP1 homologue; variable surface antigen; VSA4; rickettsia;
KW DNA vaccine.

XX OS Ehrlichia chaffeensis.

XX FH Key Location/Qualifiers

FT Peptide 1..25

FT /note= "putative signal peptide"

XX WO9816554-A1.

XX PD 23-APR-1998.

XX PF 17-OCT-1997; 97WO-US19044.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, BurrIDGE MJ, Ganta RR, Mahan SM, McGuire TC;
PI Nyika A, Rurangirwa FR;

XX DR WPI; 1998-251232/22.

XX DR N-PSDB; AAV07179.

XX Composition containing nucleic acid encoding rickettsial antigen -
PT useful for, e.g. stimulating protective immune response in humans or
PT animals

XX Claim 3; Fig 2B; 39pp; English.

XX This is the full-length variable surface antigen VSA4 protein of
CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)
CC of E. chaffeensis that was obtained on the basis of homology to the
CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
CC This genomic locus included 5 ORFs encoding similar, but
CC non-identical proteins (see AAW51091-95). A claimed composition
CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
CC (see AAW51088-99) that elicits a protective immune response against a

CC rickettsial pathogen. The nucleic acid is used, in human or
CC veterinary medicine, in vaccines to protect against Rickettsia,
CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
CC polypeptides can be used diagnostically to detect antibodies
CC associated with Ehrlichia infection (claimed).
XX

DD 60 NIIIGVFLAQDWG 74

RESULT 17
AAU04198
ID AAU04198 standard; Protein; 280 AA.

XX
XX
AC
XX
DT
AAU04198;
23-OCT-2001
(first entry)

DE
XX
KW
KW
KW

variable surface antigen 4 (VSA4) from *Ehrlichia chaffeensis*.

Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
infection; heartwater; diagnostic; variable surface antigen VSA.

XX
XX OS Ehrlichia chaffeensis.
XX
XX
PN US6251872-R1

XX
XX
PD
YV
26-JUN-2001.

PF 17-OCT-1997; 97US-0953326.
 XX
 PR 17-OCT-1996; 96US-073230.
 YY

PA (UYFL) UNIV FLORIDA.
XX
XX
PI Barbet AF, Ganta RR, McGuire TC, Burrledge MJ, Nyika A;
DI Burgescioru, EB, Maheshwari SM, Bousquet MV, Allen, AP.

XX WPI: 2001-424487/45.
DR N-PSDB; AAS07578.
XX

New MAP2 genes and polypeptides useful as vaccines for conferring immunity to human and animal rickettsial diseases, e.g. heartwater, or as molecular markers in nucleic acid analysis procedures

PS
XX
CC

CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.

the MAP polypeptides were used as nuclear matrix immunoprecipitation procedures, and to produce the MAP polypeptides, which may be used to raise antibodies that are reactive with the polypeptides. The nucleic acids may further be used as probes to identify

complementary sequences with all other nucleic acid molecules or genomes, where such probes can be applied to identify or distinguish infectious strains of organisms in diagnostic procedures or in rickettsial research where identification of particular organisms or strains is

CC	needed.	
XX		
SQ	Sequence	280 AA;

Query Match 63.9% Score 62; DB 22; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.038;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
| | : : | | | : |
Db 60 NTTIGVFGLKQDWG 74

RESULT 18
AAU73417

ID AAU73417 standard; Protein; 280 AA.
AC AAU73417;
XX
XX
DT 12-MAR-2002 (first entry)
XX
XX Ehrlichia chaffeensis outer membrane protein P28-18.
DE
XX
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX
XX Ehrlichia chaffeensis.
XX
XX WO200183699-A2.
PN
XX
XX
PD 08-NOV-2001.
XX
XX 01-MAY-2001; 2001WO-US13997.
PF
XX 01-MAY-2000; 2000US-201035P.
XX
XX (RERE-) RES DEV FOUND.
PA
XX Walker DH, Yu X;
PI
XX WPI; 2002-066527/09.
XX
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis
PT
XX Disclosure; Figure 2; 97pp; English.
XX
XX The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
XX Sequence 280 AA;
SQ
Query Match 63.9%; Score 62; DB 23; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.038;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWEG 15
| | : : ||||| : |
Db 60 NTTIGVFLKQDWG 74
RESULT 19
ABG30749
ID ABG30749 standard; Peptide; 19 AA.
XX
XX AC ABG30749;
XX
XX 21-OCT-2002 (first entry)
DT
XX Ehrlichia chaffeensis peptide fragment #5.
DE
XX Antibody detection; monoclonal antibody; polyclonal antibody.
KW
XX Ehrlichia chaffeensis.
OS
XX WO200257794-A2.
PN
XX 25-JUL-2002.
PD
XX 16-JAN-2002; 2002WO-US01395.
PF
XX 18-JAN-2001; 2001US-0765739.
PR
XX (IDEX-) IDEXX LAB INC.
PA
XX

PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX WPI; 2002-599730/64.
DR
XX
XX New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
PT
XX
XX Claim 1; Page 5; 29pp; English.
PS
XX The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. chaffeensis peptide
CC fragment used in the composition of the invention.
XX
XX Sequence 19 AA;
SQ
Query Match 60.8%; Score 59; DB 23; Length 19;
Best Local Similarity 60.0%; Pred. No. 0.0058;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWEG 15
| | : : ||||| : |
Db 1 NTTIGVFLKQDWG 15
RESULT 20
ABG30744
ID ABG30744 standard; Peptide; 20 AA.
XX
XX AC ABG30744;
XX
XX 21-OCT-2002 (first entry)
DT
XX Ehrlichia canis peptide fragment #2.
DE
XX Antibody detection; monoclonal antibody; polyclonal antibody.
KW
XX Ehrlichia canis.
OS
XX WO200257794-A2.
PN
XX 25-JUL-2002.
PD
XX 16-JAN-2002; 2002WO-US01395.
PF
XX 18-JAN-2001; 2001US-0765739.
PR
XX (IDEX-) IDEXX LAB INC.
PA
XX Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX WPI; 2002-599730/64.
DR
XX New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
PT
XX
XX Claim 1; Page 5; 29pp; English.
PS
XX The invention relates to a composition of matter comprising a polypeptide
CC

CC isolated from Ehrlichia species. The composition can be used for
 CC detecting the presence of antibodies to Ehrlichia, comprising contacting
 CC one or more polypeptides with a test sample suspected of comprising
 CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 CC complexes to form and detecting the complexes, where the detection of
 CC polypeptide/antibody complexes is an indication that the detection of
 CC Ehrlichia are present in the test sample. The composition that antibodies to
 CC detecting or quantifying in the presence of E. canis or E. chaffeensis
 CC infection in mammals, the polypeptides can be used to develop monoclonal
 CC and/or polyclonal antibodies that can be employed in assay systems and in
 CC the generation of chimeric antibodies for therapeutic use or other
 CC similar applications. This sequence represents an E. canis peptide
 CC fragment used in the composition of the invention.

XX Sequence 20 AA;

Query Match 60.8%; Score 59; DB 23; Length 20;
 Best Local Similarity 60.0%; Pred. No. 0.0062; 3; Mismatches 0; Gaps 0;
 Matches 9; Conservative

OY 1 NPTVALYGLKQDWEG 15
 DB 1 NPTTGVFGLKQDWGD 15

RESULT 21
 AAY06948
 ID AAY06948 standard; Protein; 280 AA.

AC AAY06948;

DT 05-JUL-1999 (first entry)

DE E. chaffeensis OMP-1F protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

OS Ehrlichia chaffeensis.

PN WO9913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

DR WPI; 1999-254290/21.

DR N-PSDB; AAX34748.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis

PS Claim 16; Fig 8B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, 1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 280 AA;

Query Match 60.8%; Score 59; DB 20; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.12;
 Matches 9; Conservative 3; Mismatches 0; Gaps 0;

OY 1 NPTVALYGLKQDWEG 15
 DB 60 NPTTGVFGLKQDWGD 74

RESULT 22

ABG77940

ID ABG77940 standard; Protein; 280 AA.

XX ABG77940;

XX 15-NOV-2002 (first entry)

XX Ehrlichia chaffeensis outer membrane protein (OMP) #6.

XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

XX Ehrlichia chaffeensis.

XX US2002120115-A1.

XX 29-AUG-2002.

XX 28-JAN-2002; 2002US-0059964.

XX 19-MAY-1999; 99US-0314701.

XX (RIKI/) RIKIHISA Y.

XX (OHAS/) OHASHI N.

XX Rikihisa Y, Ohashi N;

XX WPI; 2002-618954/66.

XX N-PSDB; ABS63281.

XX Isolated polynucleotide encoding an outer membrane protein of E. canis

XX or E. chaffeensis used in the diagnosis of infection -

XX Disclosure; Fig 8B; 49pp; English.

XX The invention relates to an isolated polynucleotide encoding an outer

XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used

XX in the diagnosis of infection. An infection such as human ehrlichiosis or

XX canine ehrlichiosis can be diagnosed by providing a serum sample from the

XX patient, providing a polypeptide or mixture of polypeptides, contacting

XX the sample with the polypeptide and assaying for the formation of a

XX complex between antibodies in the serum sample and the polypeptide, where

XX formation of a complex is indicative of infection with E. chaffeensis.

XX This sequence represents an Ehrlichia outer membrane protein of the

XX invention.

XX SQ Sequence 280 AA;

Query Match 60.8%; Score 59; DB 23; Length 280;

Best Local Similarity 60.0%; Pred. No. 0.12;

Matches 9; Conservative 3; Mismatches 0; Gaps 0;

OY 1 NPTVALYGLKQDWEG 15

DB 60 NPTTGVFGLKQDWGD 74

RESULT 23

AAU96110

ID AAU96110 standard; Protein; 280 AA.

XX AAU96110;

XX 02-JUL-2002 (first entry)

XX Ehrlichia chaffeensis OMP-1F.

XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;

XX KW

KW antibacterial.
 XX Ehrlichia chaffeensis.
 XX WO200222782-A2.
 PN 21-MAR-2002.
 PD 12-SEP-2001; 2001WO-US28759.
 XX 12-SEP-2000; 2000US-0660587.
 PR (RERE-) RES DEV FOUND.
 XX Walker DH, Yu X, McBride JW;
 PI WPI; 2002-351882/38.
 XX New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
 PT Example 3; Figure 3; 106pp; English.
 PS The invention relates to a recombinant homologous 28 kDa immunodominant
 XX protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AA096100-AA096118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX Sequence 280 AA;
 SQ Query Match 60.8%; Score 59; DB 23; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.12;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWEG 15
 | | :|||||:
 Db. 60 NTTGVFGLKQDWDG 74

RESULT 24
 AA06959
 ID AAY06959 standard; Protein; 288 AA.
 XX AAY06959;
 AC 05-JUL-1999 (first entry)
 XX E. canis P30 protein.
 DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX Ehrlichia canis.
 OS WO9913720-A1.
 PN 25-MAR-1999.
 PD 18-SEP-1998; 98WO-US19600.
 XX 19-SEP-1997; 97US-0059353.
 PR (OHIS) UNIV OHIO STATE.
 PA Ohashi N, Rikihisa Y;
 XX WPI: 1999-254290/21.
 DR N-PSDB; AAX34759.
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT

PT Ehrlichia canis
 XX Disclosure; Fig 19B; 55pp; English.
 PS The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX Sequence 288 AA;
 SQ Query Match 60.8%; Score 59; DB 20; Length 288;
 Best Local Similarity 60.0%; Pred. No. 0.13;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWEG 15
 | | :|||||:
 Db. 60 NTTGVFGLKQDWDG 74

RESULT 25
 ABG77950
 ID ABG77950 standard; Protein; 288 AA.
 XX ABG77950;
 AC 15-NOV-2002 (first entry)
 XX Ehrlichia canis outer membrane protein (P30F) #1.
 DE Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
 KW Ehrlichia canis.
 OS US2002120115-A1.
 PN 29-AUG-2002.
 PD 28-JAN-2002; 2002US-0059964.
 XX 19-MAY-1999; 99US-0314701.
 PR (RIKI/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.
 XX Rikihisa Y, Ohashi N;
 PI WPI: 2002-618954/66.
 DR N-PSDB; ABS63291.
 XX Isolated polynucleotide encoding an outer membrane protein of E. canis
 PT or E. chaffeensis used in the diagnosis of infection -
 PT Claim 10; Fig 19B; 49pp; English.
 PS The invention relates to an isolated polynucleotide encoding an outer
 XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.
 XX Sequence 288 AA;
 SQ Query Match 60.8%; Score 59; DB 23; Length 288;
 Best Local Similarity 60.0%; Pred. No. 0.13;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```
QY      1 NPTVALYGLKQDWEG 15
DE      | | :|:|:|:|:|
Db      60 NTTGVFGLKQDWG 74

RESULT 26
ABG30745
ID      ABG30745 standard; Peptide; 20 AA.
XX      AC
XX      ABG30745;
DT      21-OCT-2002 (first entry)
DE      Ehrlichia chaffeensis peptide fragment #1.
XX      KW
XX      Antibody detection; monoclonal antibody; polyclonal antibody.
XX      OS
XX      Ehrlichia chaffeensis.
XX      PN
XX      WO200257794-A2.
XX      PD
XX      25-JUL-2002.
XX      PF
XX      16-JAN-2002; 2002WO-US01395.
XX      PR
XX      18-JAN-2001; 2001US-0765739.
XX      PA
XX      (IDEX-) IDEXX LAB INC.
XX      PI
XX      Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX      WPI; 2002-599730/64.
XX      PT
XX      New composition of matter comprising a polypeptide, useful in detecting
XX      the presence of antibodies to Ehrlichia canis or chaffeensis, or in
XX      detecting or quantifying the presence of Ehrlichia infection in mammals
XX      PT
XX      Claim 1; Page 5; 29pp; English.
XX      CC
XX      The invention relates to a composition of matter comprising a polypeptide
XX      isolated from Ehrlichia species. The composition can be used for
XX      detecting the presence of antibodies to Ehrlichia, comprising contacting
XX      one or more polypeptides with a test sample suspected of comprising
XX      antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
XX      complexes to form and detecting the complexes, where the detection of
XX      polypeptide/antibody complexes is an indication that antibodies to
XX      Ehrlichia are present in the test sample. The composition is useful for
XX      detecting or quantifying the presence of E. canis or E. chaffeensis
XX      infection in mammals. The polypeptides can be used to develop monoclonal
XX      and/or polyclonal antibodies that can be employed in assay systems and in
XX      the generation of chimeric antibodies for therapeutic use or other
XX      similar applications. This sequence represents an E. chaffeensis peptide
XX      fragment used in the composition of the invention.
XX      SQ
XX      Sequence 20 AA;

Query Match      59.8%; Score 58; DB 23; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.0091;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 NPTVALYGLKQDWEG 15
DE      | | :|:|:|:|:|
Db      1 NTTGVFGLKQNDG 15

RESULT 27
AY06942
ID      AY06942 standard; Protein; 256 AA.
XX      AC
XX      AAY06942;
XX      ..
```

```
DT      05-JUL-1999 (first entry)
XX      E. chaffeensis p28 protein.
XX      KW
XX      Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; p30;
XX      detection; dog.
XX      OS
XX      Ehrlichia chaffeensis.
XX      PN
XX      WO9913720-A1.
XX      PD
XX      25-MAR-1999.
XX      PF
XX      18-SEP-1998; 98WO-US19600.
XX      PR
XX      19-SEP-1997; 97US-0059353.
XX      PA
XX      (OHIS ) UNIV OHIO STATE.
XX      PI
XX      Ohashi N, Rikihisa Y;
XX      WPI; 1999-254290/21.
XX      DR
XX      N-PSDB; AAX34742.
XX      PT
XX      Novel outer membrane proteins from Ehrlichia chaffeensis and
XX      Ehrlichia canis
XX      XS
XX      Claim 18; Fig 1; 55pp; English.
XX      CC
XX      The invention provides isolated outer membrane proteins (OMP) from
XX      Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX      of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX      in AAY06943-958. The E. canis proteins form part of the p30 family and
XX      consist of proteins shown in AAY06959-970. The proteins and genes are
XX      used to detect E. chaffeensis in patients and E. canis in dogs.
XX      SQ
XX      Sequence 256 AA;

Query Match      59.8%; Score 58; DB 20; Length 256;
Best Local Similarity 60.0%; Pred. No. 0.16;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 NPTVALYGLKQDWEG 15
DE      | | :|:|:|:|:|
Db      34 NTTGVFGLKQNDG 48

RESULT 28
ABG77966
ID      ABG77966 standard; Protein; 256 AA.
XX      AC
XX      ABG77966;
XX      DT
XX      15-NOV-2002 (first entry)
XX      DE
XX      Protein encoded by Ehrlichia chaffeensis p28 gene.
XX      KW
XX      Outer membrane protein; OMP; p30F; ehrlichiosis; infection; p28; OMP-1.
XX      OS
XX      Ehrlichia chaffeensis.
XX      PN
XX      US2002120115-A1.
XX      PD
XX      29-AUG-2002.
XX      PF
XX      28-JAN-2002; 2002US-0059964.
XX      PR
XX      19-MAY-1999; 99US-0314701.
XX      PA
XX      (RIKI/) RIKIHISA Y.
XX      PA
XX      (OHAS/) OHASHI N.
XX      PI
XX      Rikihisa Y, Ohashi N;
```

of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid residues. The VSA5 amino acid sequence was deduced from a partial open reading frame (ORF5) of a genomic locus (see AAV07179) of E. chaffeensis that was obtained on the basis of homology to the major antigenic protein MAP1 (see AAWS1088) of Cowdria ruminantium. This genomic locus included 5 ORFs encoding similar, but non-identical proteins (see AAWS1091-95). A claimed composition comprises a nucleic acid (see AAV07176-82) encoding a polypeptide (see AAWS1088-99) that elicits a protective immune response against a rickettsial pathogen. The nucleic acid is used, in human or veterinary medicine, in vaccines to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic polypeptides can be used diagnostically to detect antibodies associated with Ehrlichia infection (claimed).

XX
XX Sequence 276 AA;

Query Match 59.8%; Score 58; DB 19; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.18; Indels 0; Gaps
Matches 9; Conservative 4; Mismatches 2;

Qy 1 NPTVALYGLKQDWEG 15
 I I I : : I I I : I
Db 59 NTTVGVEGLKNWDG 73

RESULT 30
AAB36189
ID AAB36189 standard; Protein; 276 AA.
AC AAB36189;
XX
XX 02-MAR-2001 (first entry)
DT
DE Ehrlichia chaffeensis partial VSA5.
XX
KW Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
major antigenic protein 1; antirickettsial; vaccine; gene therapy;
Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
3gdorf3.
KW
XX
OS Ehrlichia chaffeensis.
XX
PN W0200065063-A2.
XX
XX 02-NOV-2000.
PD
PF 21-APR-2000; 200WO-US10886.
XX
PR 22-APR-1999; 99US-0130725.
XX
PA (UYFL) UNIV FLORIDA.
XX
PI Barbet AF, Bowie MV, Ganta RR, Burrridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allemen AR;
XX
XX WPI; 2000-679675/66.
DR N-PSDB; AAC68706.
XX
XX New polynucleotides useful as DNA vaccines for conferring immunity to
rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
_ .
Claim 3; Page 47; 63pp; English.

XX
XX The present sequence shows a high degree of similarity to the major
antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
used in a vaccines to protect animals or humans against rickettsial
diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
protective against the rickettsial pathogen. The nucleic acid vaccines
can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.

CC Cowdria ruminantium genes designated map 2, ihworf3, 4hworf1, 18hworf1
 CC and 3dowf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX
 SQ Sequence 276 AA;

Query Match 59.8%; Score 58; DB 21; Length 276;
 Best Local Similarity 60.0%; Pred. No. 0.18;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
 I I I : : I I I : : I
 Db 59 NTTGVGFLKQNWGD 73

RESULT 31
 AAU04199
 ID AAU04199 standard; Protein; 276 AA.
 XX
 AC AAU04199;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.
 XX
 KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN US6251872-B1.
 XX
 PI 26-JUN-2001.
 XX
 PD
 XX
 PF 17-OCT-1997; 97US-0953326.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX
 DR WPI: 2001-424487/45.
 XX
 DR N-PSDB; AAS07578.
 XX
 PT New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures -
 PS Example 3; Fig 2A-2B; 30pp; English.
 XX
 CC The sequence represents the amino acid sequence of variable surface
 CC antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.
 XX
 SQ Sequence 276 AA;

Query Match 59.8%; Score 58; DB 22; Length 276;
 Best Local Similarity 60.0%; Pred. No. 0.18;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
 I I I : : I I I : : I
 Db 59 NTTGVGFLKQNWGD 73

RESULT 32
 AAW51089
 ID AAW51089 standard; Protein; 280 AA.
 XX
 AC AAW51089;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Ehrlichia chaffeensis major antigenic protein 1 (MAP1).
 XX
 KW MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO9816554-A1.
 XX
 PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 DR WPI: 1998-251232/22.
 XX
 DR N-PSDB; AAV07177.
 XX
 PT Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 FS Claim 3; Page 18-19; 39pp; English.
 XX
 CC This polypeptide comprises the major antigen protein 1 gene (MAP1)
 CC of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see
 CC AAV07177). A claimed composition comprises a nucleic acid (see
 CC AAV07176-82) encoding a polypeptide (see AAW51089-99) that elicits a
 CC protective immune response against a rickettsial pathogen. The
 CC nucleic acid is used, in human or veterinary medicine, in vaccines
 CC to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
 CC species. The nucleic acid does not replicate in the host but
 CC remains episomal and capable of expressing polypeptide for at least
 CC 19 mth. The Ehrlichia antigenic polypeptides can be used
 CC diagnostically to detect antibodies associated with Ehrlichia
 CC infection (claimed).
 XX
 SQ Sequence 280 AA;

Query Match 59.8%; Score 58; DB 19; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.18;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWEG 15
 I I I : : I I I : : I
 Db 60 NTTGVGFLKQNWGD 74

RESULT 33
 AAB36183
 ID AAB36183 standard; Protein; 280 AA.
 XX
 AC AAB36183;
 XX

DT 02-MAR-2001 (first entry)
 XX Ehrlichia chaffeensis MAP1.
 DE
 XX Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
 KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
 KW 4hworf1; 18hworf1; 3gdorf3.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO200065063-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 21-APR-2000; 2000WO-US10886.
 XX
 PR 22-APR-1999; 99US-0130725.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 XX
 DR WPI; 2000-679675/66.
 DR N-PSDB; AAC68700.
 XX
 PT New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 PT
 XX
 PS Claim 3; Page 35-36; 63pp; English.
 XX
 CC The present sequence is given in a specification relating to nucleic
 CC acid vaccines containing genes to protect animals or humans against
 CC rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
 CC sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The vaccine comprises the
 CC major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
 CC (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
 CC driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
 CC ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1 and
 CC 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX
 SQ Sequence 280 AA;
 Query Match 59.8%; Score 58; DB 21; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.18;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWEG 15
 | | | : | | | : | |
 Db 60 NTTGVGFLKQKQNDG 74
 RESULT 34
 AAU04193
 ID AAU04193 standard; Protein; 280 AA.
 XX
 AC AAU04193;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.
 XX
 KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PI

PN US6251872-B1.
 XX
 PD 26-JUN-2001.
 XX
 PF 17-OCT-1997; 97US-0953326.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX
 DR WPI; 2001-424487/45.
 DR N-PSDB; AAS07576.
 XX
 PT New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures
 XX
 PS Disclosure; Column 15-17; 30pp; English.
 XX
 CC The sequence represents the amino acid sequence of major antigenic
 CC protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
 CC polypeptides are useful as vaccines for conferring immunity to rickettsia
 CC infection, including Cowdria ruminantium causing heartwater. The MAP
 CC polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.
 XX
 SQ Sequence 280 AA;
 Query Match 59.8%; Score 58; DB 22; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.18;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWEG 15
 | | | : | | | : | |
 Db 60 NTTGVGFLKQKQNDG 74
 RESULT 35
 AAY06943
 ID AAY06943 standard; Protein; 281 AA.
 XX
 AC AAY06943;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 DE E. chaffeensis OMP-1 protein.
 XX
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO9913720-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 18-SEP-1998; 98WO-US19600.
 XX
 PR 19-SEP-1997; 97US-0059353.
 XX
 PA (OHIS) UNIV OHIO STATE.
 XX
 PI Ohashi N, Rikihisa Y;

```

XX WPI: 1999-254290/21.
DR N-PSDB; AAX34743.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
XX Disclosure: Fig 3B; 55pp; English.
PS
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1 (P1 to P2) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX SQ Sequence 281 AA;
SQ
Query Match 59.8%; Score 58; DB 20; Length 281;
Best Local Similarity 60.0%; Pred. NO. 0.18;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWEG 15
Db 59 NTTGVGFLKQNWIDG 73

```

Search completed: October 6, 2003, 07:48:01
Job time : 32.8358 secs


```

Y33B_MYCPN
ID Y33B_MYCPN STANDARD; PRT; 341 AA.
AC P75302;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG335.2 homolog (P01_orf341).
GN MPN483 OR MP359
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC STRONG, TO M.GENITALIUM MG335.2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000034; BAB96007.1; -
CC FIC: S73685; S73685.
CC InterPro: IPR001173; Glyco_trans_2.
CC Pfam: PF00535; Glycos_transf_2; 1.
CC Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 341 AA; 40414 MW; C209F50D714CB3D0 CRC64;

Query Match 48.0%; Score 47; DB 1; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 LKQDNGV 16
DB 26 LKQDNGV 33
|||||

RESULT 2
YB_JF_SALTY
ID YB_JF_SALTY STANDARD; PRT; 375 AA.
AC Q82841;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical RNA methyltransferase YbJf (EC 2.1.1.-).
YB_JF OR SYR0915 OR Y2014.
GN Salmonella typhi.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.,

"Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
Nature 413:848-852(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanani V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar typhi strains Ty2
and CT18.";
J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Could be a 23S rRNA (Uracil-5-)-methyltransferase.
CC -!- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. YB_JF
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL627268; CAD05321.1; -
CC EMBL: AE016840; AAO69626.1; -
CC HAMAP: MF_01012; -; 1
CC InterPro: IPR000051; SAM_bind.
CC InterPro: IPR001566; TrmA.
CC PROSITE: PS01230; TrmA_1; 1.
CC PROSITE: PS01231; TrmA_2; 1.
CC KW Hypothetical protein; Transferase; Methyltransferase; rRNA processing;
KW Complete proteome.
FT ACT_SITE 334 BY SIMILARITY.
SQ SEQUENCE 375 AA; 42070 MW; 40C6A0563FA92598 CRC64;

Query Match 45.9%; Score 45; DB 1; Length 375;
Best Local Similarity 58.8%; Pred. No. 4.3;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 1 NPTVA--LYGLKQDNG 15
DB 214 NPTVASRLYATARDVWG 230
|||||

RESULT 3
YB_JF_SALTY
ID YB_JF_SALTY STANDARD; PRT; 376 AA.
AC Q82035;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Hypothetical RNA methyltransferase YbJf (EC 2.1.1.-).
YB_JF OR STM0882.
GN Salmonella typhimurium.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Could be a 23S rRNA (Uracil-5-)-methyltransferase.
CC -!- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. YB_JF
CC SUBFAMILY.

```

```

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AE008737; AAL19818.1; -
StyGene; SG22222; ybjf.
HAWAP; ME_01012; -; 1.
InterPro; IPR000051; SAM_bind.
InterPro; IPR001566; TRMA.
DR PROSITE; PS01230; TRMA_1; 1.
DR PROSITE; PS01231; TRMA_2; 1.
KW Hypothetical protein; Transferase; Methyltransferase; rRNA processing;
KW Complete proteome. 334 BY SIMILARITY.
FT ACCT_SITE 334 334 334 334 334 334 334 334 334 334 334 334 334 334 334
SQ SEQUENCE 376 AA; 42180 MW; 343163D8E7FE0BFF CRC64;
Query Match 45.9%; Score 45; DB 1; Length 376;
Best Local Similarity 58.8%; Pred. No. 4.4;
Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;
Oy 1 NPTVA--LYGLKQDWNG 15
||||| ||| |||
Db 214 NPTVASRLYATARDWVG 230
RESULT 4
MTHC_DROME STANDARD; PRT; 475 AA.
AC P83119;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable G-protein-coupled receptor Mth-like 12 precursor (Methuselah-
like 12 protein).
GN MTHL12 OR MTH-LIKE-12.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

```

```

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP IDENTIFICATION.
RX MEDLINE=21173629; PubMed=11274391;
RA West A.P. Jr., Llamas L.L., Snow P.M., Benzer S., Bjorkman P.J.;
RT "Crystal structure of the ectodomain of Methuselah, a Drosophila G
RT protein-coupled receptor associated with extended lifespan.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3744-3749(2001).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED
CC RECEPTORS. MTH SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: AE003699; -; NOT_ANNOTATED_CDS.
DR FlyBase; FBgn0045442; mthl12.
DR GO; GO:0004930; F:G-protein coupled receptor activity; ISS.
DR GO; GO:0008340; P:determination of adult life span; ISS.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; ISS.
DR GO; GO:0006950; P:response to stress; ISS.
DR InterPro; IPR000832; GPCR_secretin.
DR Pfam; PF00002; 7tm2; 1.
DR PROSITE; PS00649; G-PROTEIN_RECEP_F2_1; FALSE_NEG.
DR PROSITE; PS00650; G-PROTEIN_RECEP_F2_2; FALSE_NEG.
DR PROSITE; PS0261; G-PROTEIN_RECEP_F2_4; 1.
KW Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 17
FT CHAIN 18 475
FT PROBABLE G-PROTEIN-COUPLED RECEPTOR MTH-
FT LIKE 12.
FT DOMAIN 18 214
FT TRANSMEM 215 235
FT DOMAIN 236 242
FT TRANSMEM 243 263
FT DOMAIN 264 275
FT TRANSMEM 276 296
FT DOMAIN 297 307
FT TRANSMEM 308 328
FT DOMAIN 329 360
FT TRANSMEM 361 381
FT DOMAIN 382 403
FT TRANSMEM 404 424
FT DOMAIN 425 442
FT TRANSMEM 443 463
FT DOMAIN 464 475
FT DISULFID 27 81
FT DISULFID 83 88
FT DISULFID 92 183
FT DISULFID 93 104
FT DISULFID 149 203
FT CARBOHYD 19 19
FT CARBOHYD 34 34
FT CARBOHYD 55 55
FT CARBOHYD 135 135
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT N-LINKED (GLCNAC. .) (POTENTIAL).

```

```
FT CARBOHYD 352 352 N-LINKED (GLCNAC...)(POTENTIAL).
SQ SEQUENCE 475 AA: 55424 MW: 3590FEF875F77264 CRC64;
Query Match 43.9%; Score 43; DB 1; Length 475;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDNGVVS 17
Db 340 PLVYFGLFTDWNSSS 355

RESULT 5
QALIS_NEUCR
ID QALIS_NEUCR STANDARD; PRT; 918 AA.
AC P11637;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Quinate repressor.
GN QAL-1S.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RC MEDLINE=89293848; PubMed=2525625;
RX Geever R.E., Huilet L., Baum J.A., Tyler B.M., Patel V.B.,
RA Rutledge B.J., Case M.E., Giles N.H.;
RT "DNA sequence, organization and regulation of the qa gene cluster of
RT Neurospora crassa.";
RL J. Mol. Biol. 207:15-34(1989).
CC -!- FUNCTION: REPRESSOR FOR ENZYMES AND PROTEINS OF QUINATE
CC METABOLISM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14603; CAA32753.1; -.
CC PIR: S04255; S04255.
CC InterPro: IPR001381; DHQuinase_I.
CC InterPro: IPR000623; Shik_kinase.
CC InterPro: IPR006151; Shikimate_DH.
CC Pfam: PF01487; DHQuinase_I.
CC Pfam: PF01488; Shikimate_DH; 1.
CC Pfam: PF01202; SKI; 1.
CC Quinate metabolism; Transcription regulation; Repressor; DNA-binding.
KW Quinate metabolism; 100580 MW; 67EDR399CBF098B2 CRC64;
SQ SEQUENCE 918 AA: 100580 MW: 67EDR399CBF098B2 CRC64;

Query Match 43.9%; Score 43; DB 1; Length 918;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 ALYGLKQDNGVSA 18
Db 684 ALYGTIDWIGIRA 697

RESULT 6
YJL3_YEAST
ID YJL3_YEAST STANDARD; PRT; 1803 AA.
AC P47024; P87192;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Transposon Ty4 207.7 kDa hypothetical protein.
DE Transposon Ty4 207.7 kDa hypothetical protein.
```

```
GN TY4B OR YJL113W OR J0780.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RC MEDLINE=97103775; PubMed=8948101;
RX Cziepluch C., Kordes E., Pujol A., Jauniaux J.-C.;
RA "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X
RA reveals 19 open reading frames including URA2 (5' end), TRK1, PBS2,
RT SPT10, GCD14, RPE1, PHO86, NCA3, ASF1, CCT7, GZF3, two tRNA genes,
RT three remnant delta elements and a Ty4 transposon.";
RL Yeast 12:1471-1474(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z49389; CAA89409.1; -.
CC SGD: S0003649; YJL113W.
CC InterPro: IPR001584; Rve.
CC InterPro: IPR001878; Znf_CCHC.
CC Pfam: PF00665; rve; 1.
CC SMART: SM00343; Znf_C2HC; 1.
CC Transposable element; Hypothetical protein.
KW Transposable element; 207693 MW; 16DCD7284A8D52D3 CRC64;
SQ SEQUENCE 1803 AA: 207693 MW; 16DCD7284A8D52D3 CRC64;

Query Match 43.4%; Score 42.5; DB 1; Length 1803;
Best Local Similarity 69.2%; Pred. No. 60;
Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 5 ALYGLKQ---DWN-14
Db 1415 ALYGLKQSPKEWN 1427

RESULT 7
RT29_BOVIN
ID RT29_BOVIN STANDARD; PRT; 122 AA.
AC P82922;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mitochondrial 26S ribosomal protein S29 (S29mt) (MRP-S29) (Fragments).
GN DAP3 OR MRPS29.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=21276436; PubMed=11279123;
RA Koc E.C., Burkhardt W., Blackburn K., Moseley A., Spremulli L.L.;
RT "The small subunit of the mammalian mitochondrial ribosome:
RT identification of the full complement of ribosomal proteins present.";
RL J. Biol. Chem. 276:19363-19374(2001).
CC -!- SUBUNIT: Component of the mitochondrial ribosome small subunit
CC (28S) which comprises a 12S rRNA and about 30 distinct proteins.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC Ribosomal protein; Mitochondrion.
KW NON_TER 1
FT NON_TER 34
FT NON_CONS 51
FT NON_CONS 52
FT NON_CONS 67
FT NON_CONS 68
FT NON_CONS 76
FT NON_CONS 77
```

FT NON_CONS 88 89
 FT NON_CONS 100 101
 FT NON_CONS 110 111
 FT NON_TER 122 122
 SQ SEQUENCE 122 AA; 14053 MW; 5D2CCAEAREDA14D26 CRC64;

Query Match 42.9%; Score 42; DB 1; Length 122;

Best Local Similarity 60.0%; Pred. No. 4.3;

Matches 9; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

Qy 1 NPTV--ALYGLKQDW 13

Db 40 HPAVRYLYGKQDW 54

RESULT 8

CDX2_HUMAN STANDARD; PRT; 313 AA.

ID Q99626; O00503; Q969L8;

AC 01-NOV-1997 (Rel. 35, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Homeobox protein CDX-2 (Caudal-type homeobox protein 2) (CDX-3).

GN CDX2 OR CDX3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-Colon carcinoma;

RX MEDLINE=97188282; PubMed=9036867;

RA Mallo G.V., Rechreche H., Frigerio J.M., Rocha D., Zweibaum A.,

RA Lacasa M., Jordan B.R., Dusetti N.J., Dagorn J.C., Iovanna J.L.,

RT "Molecular cloning, sequencing and expression of the mRNA encoding

human Cdx1 and Cdx2 homeobox. Down-regulation of Cdx1 and Cdx2 mRNA

expression during colorectal carcinogenesis.";

RL Int. J. Cancer 74:35-44(1997).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=98120622; PubMed=9459001;

RA Drummond F.J., Pütt W., Fox M., Edwards Y.H.;

RT "Cloning and chromosome assignment of the human CDX2 gene.";

RL Ann. Hum. Genet. 61:393-400(1997).

RN [3]

RP SEQUENCE FROM N.A.

RA Tanizawa Y., Ueda K., Inoue H., Ayame H., Aoki M., Kuwano A.,

RA German M.S., Liu L., Donis-Keller H., Permutt M.A., Oka Y.;

RT "Isolation, characterization, and linkage mapping of the human

caudal-type homeobox gene, CDX2/3.";

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=21094877; PubMed=11161380;

RA Sivagnanasundaram S., Islam I., Talbot I., Drummond F., Walters J.R.,

RA Edwards Y.H.;

RT "The homeobox gene CDX2 in colorectal carcinoma: a genetic analysis.";

RL Br. J. Cancer 84:218-225(2001).

RN [5]

RP SEQUENCE FROM N.A.

RC TISSUE-Colon;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausberg R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madañ A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madañ A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: Involved in the transcriptional regulation of multiple
 CC genes expressed in the intestinal epithelium. Important in broad
 CC range of functions from early differentiation to maintenance of
 CC the intestinal epithelial lining of both the small and large
 CC intestine.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- PTM: Phosphorylation of Ser-60 mediates the transactivation
 CC capacity (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE CAUDAL HOMEBOX FAMILY.
 CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CDX2ID326.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch)
 CC -----

DR EMBL; U51096; AAB40603.1; -;

DR EMBL; Y13709; CAA74038.1; -;

DR EMBL; AF007886; AAD05200.1; -;

DR EMBL; AF007884; AAD05200.1; JOINED.

DR EMBL; AF007885; AAD05200.1; JOINED.

DR EMBL; AJ278431; CAB94779.1; -;

DR EMBL; AJ278432; CAB94779.1; JOINED.

DR EMBL; AJ278434; CAB94779.1; JOINED.

DR EMBL; BC014461; AAH14461.1; -;

DR HSSP; P14653; 1B72.

DR TRANSFAC; T03246; -;

DR Genew; HGNC:1806; CDX2.

DR MIM; 600297; -;

DR GO; GO:0003700; F:transcription factor activity; TAS.

DR GO; GO:0007397; P:histogenesis and organogenesis; TAS.

DR GO; GO:0007048; P:oncogenesis; TAS.

DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.

DR InterPro; IPR006820; Caudal_act.

DR InterPro; IPR001356; Homeobox.

DR InterPro; IPR000047; HTH_lambrepresr.

DR Pfam; PF04731; Caudal_act; 1.

DR Pfam; PF00046; homeobox; 1.

DR PRINTS; PR00024; HOMEBOX.

DR PRINTS; PR00031; HTHREPRESSR.

DR ProDom; PD000010; Homeobox; 1.

DR SMART; SM00389; HOX; 1.

DR PROSITE; PS00027; HOMEBOX_1; 1.

DR PROSITE; PS00071; HOMEBOX_2; 1.

KW Transcription regulation; Activator; Developmental protein; Homeobox;

KW DNA-binding; Nuclear protein; Phosphorylation; Polymorphism.

FT DNA_BIND 186 245 HOMEBOX.

FT DOMAIN 46 53 POLY-ALA.

FT DOMAIN 85 92 POLY-ALA.

FT DOMAIN 250 256 POLY-GLN.

FT DOMAIN 257 270 POLY-PRO.

FT MOD_RES 60 60 PHOSPHORYLATION (BY SIMILARITY).

FT VARIANT 293 293 S -> P (IN dSNP:1805107).

FT FTID=VAR_014530.

FT CONFLICT 52 53 AA -> Q (IN REF. 1).

FT CONFLICT 89 89 MISSING (IN REF. 1).

FT CONFLICT 95 95 G -> A (IN REF. 1).

SQ SEQUENCE 313 AA; 33509 MW; B5724F1F7EC8F9AF CRC64;

Query Match 42.9%; Score 42; DB 1; Length 313;

```

Best Local Similarity 61.5%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 5 ALYG--LKODWNG 15
Db 67 AAYGAPLREDWNG 79

RESULT 9
S123_PSEAM
ID S123_PSEAM STANDARD; PRT; 1023 AA.
AC P5019;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Solute carrier family 12 member 3 (Thiazide-sensitive sodium-chloride
DE cotransporter) (Na-Cl symporter).
GN SLC12A3 OR TSC.
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes
OS americanus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Bladder urothelium;
RX MEDLINE=93219361; PubMed=8464884;
RA Gamba G., Saltzberg S.N., Lombardi M., Miyashita A., Lytton J.,
RA Hediger M.A., Brenner M.B., Hebert S.C.;
RT "Primary structure and functional expression of a cDNA encoding the
RT thiazide-sensitive, electroneutral sodium-chloride cotransporter.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:2749-2753(1993).
CC -!- FUNCTION: Electrically silent transporter system. Mediates sodium
CC and chloride reabsorption.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Urinary bladder.
CC -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L11615; AAA49272.1; -
CC PIR; A47296; A47296
CC InterPro; IPR002293; AA/rel_permease1.
CC InterPro; IPR004842; KCL_cotranspt.
CC InterPro; IPR004841; Permease.
CC Pfam; PF00324; aa_permeases; 1.
CC TIGRFAMs; TIGR00930; 2a30; 1.
KW Transport; Ion transport; Sodium transport; Symport;
KW Transmembrane.
FT DOMAIN 1 132 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 133 153 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT DOMAIN 185 215 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 259 279 POTENTIAL.
FT DOMAIN 280 283 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 284 304 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT DOMAIN 358 374 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 452 472 POTENTIAL.
FT DOMAIN 473 510 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 511 531 POTENTIAL.
FT TRANSMEM 577 597 POTENTIAL.
FT DOMAIN 598 745 CYTOPLASMIC (POTENTIAL).

Query Match 42.9%; Score 42; DB 1; Length 1023;
Best Local Similarity 58.3%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDW 13
Db 727 PNVLLMGFKDW 738

RESULT 10
VP50_BPAPS STANDARD; PRT; 184 AA.
ID VP50_BPAPS
AC Q9T1P8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative protein P50.
GN 50.
OS Bacteriophage APSE-1.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
OX NCBI_TaxID=106199;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99420383; PubMed=10489345;
RA van der Wilk F., Dulleman A.M., Verbeek M., van den Heuvel J.F.J.M.;
RT "Isolation and characterization of APSE-1, a bacteriophage infecting
RT the secondary endosymbiont of acyrthosiphon pisum.";
RL Virology 262:104-113(1999).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF157835; AAF03993.1; -
CC Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 184 AA; 20416 MW; EB705FA272B6768F CRC64;

Query Match 41.8%; Score 41; DB 1; Length 184;
Best Local Similarity 50.0%; Pred. No. 9.9;
Matches 9; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 1 NPTVALYGLKODWNGVSA 18
Db 128 NATVEFYGDNNKGVSVA 145

RESULT 11
AROE_SYNY3 STANDARD; PRT; 290 AA.
ID AROE_SYNY3
AC P74591;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Shikimate 5-dehydrogenase (EC 1.1.1.25).
GN AROE OR SLR1559.
OS Synechocystis sp. (strain PCC 6803).

```

```

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyauchi N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: Shikimate + NADP(+) = 5-dehydroshikimate +
CC NADPH.
CC -!- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
CC fourth step.
CC -!- SIMILARITY: Belongs to the shikimate dehydrogenase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90916; BAA18699.1; -
DR PIR; S76787; S76787.
DR HAMAP; MF_00222; -; 1.
DR InterPro; IPR006152; Shikimate.
DR InterPro; IPR006151; Shikimate_DH.
DR Pfam; PF01488; Shikimate_DH; 1.
DR TIGRFAMs; TIGR00507; aroE; 1.
KW Aromatic amino acid biosynthesis; Oxidoreductase; NADP;
KW Complete proteome.
SQ SEQUENCE 290 AA; 31099 MW; 8A2D38EE5D57B303 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 290;
Best Local Similarity 61.5%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 6 LYGLKQDWNGVSA 18
I | | | | | | | |
Db 116 LLELKQDWNGRTA 128

RESULT 12
ALR3_SALTI
ID ALR3_SALTI STANDARD; PRT; 368 AA.
AC Q82300;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Alanine racemase 3 (EC 5.1.1.1).
GN ALR3 OR STY3763 OR T3513.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis N., Davies R.M., White N., Farrar J.,
RA Fellwell T., Hamlin P., Haque A., Hien T.T., Holroyd S., Jagers K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;

```

```

RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RX Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Provides the D-alanine required for cell wall
CC biosynthesis (By similarity).
CC -!- CATALYTIC ACTIVITY: L-alanine = D-alanine.
CC -!- COFACTOR: Pyridoxal phosphate (By similarity).
CC -!- PATHWAY: D-alanine branch of peptidoglycan biosynthesis; first
CC step.
CC -!- SIMILARITY: Belongs to the alanine racemase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL627279; CAD09518.1; -
DR EMBL; AE016846; AAO1021.1; -
DR HAMAP; MF_01201; -; 1.
DR InterPro; IPR000821; Ala_racemase.
DR Pfam; PF00842; Ala_racemase_C; 1.
DR Pfam; PF01168; Ala_racemase_N; 1.
DR PRINTS; PR00992; ALARACEMASE.
DR PROSITE; PS00395; ALANINE_RACEMASE; 1.
KW Isomerase; Pyridoxal phosphate; Cell wall; Peptidoglycan synthesis;
KW Complete proteome.
FT ACT_SITE 42 42 CATALYTIC BASE SPECIFIC TO D-ALANINE
FT ACT_SITE 262 262 (BY SIMILARITY).
FT ACT_SITE 42 42 CATALYTIC BASE SPECIFIC TO L-ALANINE
FT BINDING 42 42 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 368 AA; 39778 MW; 1338A6D2936004B3 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 368;
Best Local Similarity 58.3%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PTVALYGLKQDW 13
I | | | | | | | |
Db 222 PGVALFGVAQWP 233

RESULT 13
Y030_NPVOP
ID Y030_NPVOP STANDARD; PRT; 459 AA.
AC O10293;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 52.7 kDa protein (ORF38).
OS Orgyia pseudotsugata multicausid polyhedrosis virus (OpMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RX Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RX Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RT polyhedrosis virus genome.";

```

```

RL Virology 229:381-399(1997)).
CC -!- SIMILARITY: TO CORRESPONDING ORF IN ACMPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U75930; AAC59037.1; -.
KW Hypothetical protein.
KW SEQUENCE 459 AA; 52731 MW; DAAFBDB8C662BF596 CRC64;
Query Match 41.8%; Score 41; DB 1; Length 459;
Best Local Similarity 53.3%; Pred. NO. 26;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 4 VALYGLKQDWNGVSA 18
DB 56 LSAFGDKIDWNEVSA 70
::: |||||
RESULT 14
ID PYRG_AZOBR STANDARD; PRT; 544 AA.
AC P28595;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CTP synthase (EC 6.3.4.2) (UTP-ammonia ligase) (CTP synthetase).
GN PYRG
OS Azospirillum brasilense.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Rhodospirillaceae; Azospirillum.
OX NCBI_TaxID=192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP7 / ATCC 29145;
RC MEDLINE=941186025; PubMed=8138139;
RA Zimmer W., Hundschaen B.;
RT "Identification and sequencing of pyrG, the CTP synthetase gene of
RT Azospirillum brasilense Sp7.";
RL FEMS Microbiol. Lett. 115:273-278(1994).
CC -!- FUNCTION: Catalyzes the ATP-dependent amination of UTP to CTP with
CC either L-glutamine or ammonia as the source of nitrogen.
CC -!- CATALYTIC ACTIVITY: ATP + UTP + NH(3) = ADP + phosphate + CTP.
CC -!- ENZYME REGULATION: Allosterically activated by Gmp, when glutamine
CC is the substrate. Inhibited by Ctp.
CC -!- PATHWAY: Pyrimidine biosynthesis; conversion of UMP to CTP; third
CC (last) step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the CTP synthase family.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X67216; CAA47656.1; -.
DR PIR: I39496; S25101.
DR HAMAP: MF_01227; -.
DR InterPro: IPR000991; Gataase_1.
DR Pfam: PF00117; Gataase; 1.
DR TIGRFAMs: TIGR00337; PyrG; 1.
DR PROSITE: PS00442; GATASE_TYPE_I; 1.
KW Pyrimidine biosynthesis; Ligase; Glutamine amidotransferase.

```

```

FT DOMAIN 1 300 AMINATOR DOMAIN.
FT ACT_SITE 301 544 GLUTAMINE AMIDOTRANSFERASE.
FT ACT_SITE 380 380 GATASE (BY SIMILARITY).
FT ACT_SITE 516 516 GATASE (BY SIMILARITY).
FT ACT_SITE 518 518 GATASE (BY SIMILARITY).
SQ SEQUENCE 544 AA; 59908 MW; 13A1A8B8B6C5DB CRC64;
Query Match 41.8%; Score 41; DB 1; Length 544;
Best Local Similarity 53.8%; Pred. NO. 31;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDW 13
DB 410 NPVVGLGLMTW 422
::: |||||
RESULT 15
ID PLO3_MOUSE STANDARD; PRT; 741 AA.
AC Q9R0E1; Q9CY99;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 precursor
DE (EC 1.14.11.4) (Lysyl hydroxylase 3) (LH3).
GN PLOD3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
RX MEDLINE=99357020; PubMed=10429951;
RX Ruetalaainen H., Sipila L., Kerkela E., Pospiech H., Myllylae R.;
RT "Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3,
RT their phylogenetic analysis and tissue-specific expression in the
RT mouse.";
RL Matrix Biol. 18:325-329(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RC MEDLINE=21233587; PubMed=11334715;
RA Ruotsalainen H., Vanhatupa S., Tampio M., Sipila L., Valtavaara M.,
RA Myllylae R.;
RT "Complete genomic structure of mouse lysyl hydroxylase 2 and lysyl
RT hydroxylase 3/collagen glucosyltransferase.";
RN Matrix Biol. 20:137-146(2001).
RP [3]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Hono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: FORMS HYDROXYLYSINE RESIDUES IN -XAA-LYS-GLY- SEQUENCES
CC IN COLLAGENS. THESE HYDROXYLYSINES SERVE AS SITES OF ATTACHMENT

```

FOR CARBOHYDRATE UNITS AND ARE ESSENTIAL FOR THE STABILITY OF THE INTERMOLECULAR COLLAGEN CROSSLINKS.

-!- CATALYTIC ACTIVITY: Procollagen L-lysine + 2-oxoglutarate + O(2) = procollagen 5-hydroxy-L-lysine + succinate + CO(2).

-!- COFACTOR: IRON AND ASCORBATE.

-!- SUBUNIT: Homodimer (By similarity).

-!- SUBCELLULAR LOCATION: MEMBRANE BOUND IN CISTERNAE OF ROUGH ENDOPLASMIC RETICULUM.

-!- TISSUE SPECIFICITY: Highly expressed in the heart, lung, liver and testis.

-!- SIMILARITY: BELONGS TO THE LYSYL HYDROXYLASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF046783; AAD54618.1; -
 EMBL; AY014830; AAK00576.1; -
 EMBL; AK013195; BAB28704.1; -
 MGD; MGI:1347008; Plod3.
 InterPro; IPR005123; 2OG-FeII_Oxy.
 InterPro; IPR006620; Pro_4_hyd_alph.
 Pfam; PF03171; 2OG-FeII_Oxy; 1.
 ProDom; PD011578; ProcollLys_dioxy; 1.
 SMART; SM00702; P4Hc; 1.
 PROSITE; PS01325; LYS_HYDROXYLASE; 1.
 Oxidoreductase; Dioxxygenase; Signal; Iron; Vitamin C; Glycoprotein;
 Endoplasmic reticulum; Membrane; Polymorphism.
 SIGNAL 1 27
 CHAIN 28 741
 FT METAL 670
 FT METAL 672
 FT METAL 722
 FT ACT_SITE 732
 FT CARBOHYD 66
 FT CARBOHYD 286
 FT CARBOHYD 551
 FT CONFLICT 8
 SQ SEQUENCE 741 AA; 84922 MW; D1B79B386339D9F4 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 741;
 Best Local Similarity 53.38; Pred. No. 43;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Oy 1 NPTVALXGLKQDWNG 15
 I I I I I I I I I I
 Db 66 NYTVRTLGLGQEWNG 80

RESULT 16
 ID SVV_STRCO STANDARD; PRT; 874 AA.
 AC O06851; O9LIG3;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Valyl-trna synthetase (EC 6.1.1.9) (Valine--trna ligase) (ValRS).
 GN VALS OR SCO2615 OR SCC88.26C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S., Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A.;

"Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).";
 Nature 417:141-147(2002).

[2]
 RN SEQUENCE OF 616-873 FROM N.A.
 RP STRAIN-A3(2) / M145;
 RX MEDLINE=98164371; PubMed=9503623;
 RA Burger A., Brandt B., Suesstrunk U., Thompson C.J., Wohlleben W.;

"Analysis of a Streptomyces coelicolor A3(2) locus containing the nucleoside diphosphate kinase (ndk) and folypolyglutamate synthetase (folC) genes.";
 FEMS Microbiol. Lett. 159:283-291(1998).

RL -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate + L-valyl-tRNA(Val).
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 CC EMBL; AL939113; CAB75396.1; -
 CC EMBL; Y13070; CAA73510.1; -
 CC HSSP; P96142; IGAX.
 CC InterPro; IPR002300; tRNA-synt_1a.
 CC InterPro; IPR001412; tRNA-synt_1.
 CC InterPro; IPR002303; tRNA-synt_val.
 CC Pfam; PF00133; tRNA-synt_1; 1.
 CC PRINTS; PR00986; TRNASYNTHVAL.
 CC TIGRFAMs; TIGR00422; vals; 1.
 CC PROSITE; PS00178; AA-TRNA_LIGASE_I; 1.
 CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 57 67 "HIGH" REGION.
 FT SITE 531 535 "KMSKS" REGION.
 FT BINDING 534 534 ATP (BY SIMILARITY).
 FT CONFLICT 638 639 KL -> NV (IN REF. 1).
 SQ SEQUENCE 874 AA; 97566 MW; AD11E4871786BAA0 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 874;
 Best Local Similarity 53.8%; Pred. No. 51;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 3 TVALYGLKQDWNG 15
 I I I I I I I I I I
 Db 517 TIALHGMVRDQNG 529

RESULT 17
 PKHD_HUMAN
 ID PKHD_HUMAN STANDARD; PRT; 4074 AA.
 AC O8VCZ9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Polycystic kidney and hepatic disease 1 precursor (Fibrocystin)
 DE (Polycystin) (Tigmin).
 GN PKHD1 OR FCYT OR TIGM1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;


```

FT CARBOHYD 1460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1475 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1494 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1527 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1582 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1602 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1764 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1779 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1879 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1883 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1919 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1945 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1959 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2034 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2115 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2350 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2435 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2509 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2535 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2595 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2753 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2768 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3008 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3707 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3726 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 3838 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 3386 GTFREEQKCTY -> VTTEQALKISE (in isoform 2).
FT VARSPLIC 3397 2).
FT VARSPLIC 3397 4074 /TIG-VSP_003947.
Query Match 41.8%; Score 41; DB 1; Length 4074;
Best Local Similarity 54.5%; Pred. No. 2.5e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 ALYGLKQDWNG 15
|::| |::|
Db 2282 AIHGRKDWNG 2292

RESULT 18
CARA_METJA STANDARD; PRT; 354 AA.
AC Q58425;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase small chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase glutamine chain).
GN CARA OR MJ1019.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2561 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

```

```

RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.E., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O -> 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carA family.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67544; AAB99021.1; -.
CC HSP; P00907; LCS0.
CC TIGR; MJ1019; -.
CC HAMAP; MF_01209; -.
CC InterPro; IPR006220; Anth_synthII.
CC InterPro; IPR006274; CarA_small.
CC InterPro; IPR001317; CPS_Gatase.
CC InterPro; IPR002474; CPSase_sm_chain.
CC InterPro; IPR000991; Gatase_1.
CC Pfam; PF00988; CPSase_sm_chain; 1.
CC Pfam; PF00117; Gatase; 1.
CC PRINTS; PR00097; ANTSNTHASEII.
CC PRINTS; PR00099; CPSGATASE.
CC PRINTS; PR00096; GATASE.
CC TIGRFAMS; TIGR01368; CPSaseIismall; 1.
CC PROSITE; PS00442; GATASE_TYPE_I; FALSE_NEG.
CC Arginine biosynthesis; Pyrimidine biosynthesis; Ligase;
CC Glutamine amidotransferase; Complete proteome.
CC KW
CC DOMAIN 1 171 CFSASE.
CC FT DOMAIN 172 354 GLUTAMINE AMIDOTRANSFERASE.
CC FT ACT_SITE 246 246 GATASE (BY SIMILARITY).
CC SQ SEQUENCE 354 AA; 39920 MW; 4D8F0776CA65F9F CRG64;

Query Match 41.3%; Score 40.5; DB 1; Length 354;
Best Local Similarity 40.0%; Pred. No. 24;
Matches 8; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 2 PTVALYGLKQDW---NGVSA 18
|::| |::|
Db 56 PLEGNYGVKKDFESDGIKA 75

RESULT 19
MEGL_VIBCH STANDARD; PRT; 162 AA.
ID MEGL_VIBCH
AC Q9KFK1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-adenosylmethionine:2-demethylmenaquinone methyltransferase 1
DE (EC 2.1.1.-).
GN MENGI OR VC2366.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

```

```

OC Vibrionaceae: Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL for M16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Mayday L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
RC -!- FUNCTION: CONVERTS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK)
CC (By similarity).
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE MENQ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF004307; AAF95509.1;
DR F1R; F82084; F82084.
DR TIGR; VC2366;
DR HAMAP; MF_00471; 1.
DR InterPro; IPR005493; Methyltransf_6.
DR Pfam; PF03737; Methyltransf_6; 1.
DR Menquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
KW SEQUENCE 162 AA; 17848 MW; D28979C60AA56031 CRC64;
Query Match 40.8%; Score 40; DB 1; Length 162;
Best Local Similarity 46.2%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 4 VALYGLKQDWNGV 16
DB 79 LAIAIKNDWEGV 91
:|: :|:|:|
RESULT 20
PYRF STRMU
ID PYRF STRMU STANDARD; PRT; 230 AA.
AC Q8DTV1;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)
DE (OMPCase) (Ompdecase).
DE PYRF OR SMU.1222.
GN Streptococcus mutans.
OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carlson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RA "Genome sequence of Streptococcus mutans UAI159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate = UMP + CO(2).

```

```

CC -!- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY. SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE014959; AAN58907.1;
DR HAMAP; MF_01200; 1.
DR InterPro; IPR001754; OMPdecase.
DR Pfam; PF00215; OMPdecase; 1.
DR PROSITE; PS00156; OMPDECASE; 1.
KW Lyase; Decarboxylase; Pyrimidine biosynthesis; Complete proteome.
FT ACT_SITE 63 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 230 AA; 24915 MW; 2EC602717AB18257 CRC64;
Query Match 40.8%; Score 40; DB 1; Length 230;
Best Local Similarity 40.0%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWNG 15
DB 216 DFIAAYQAIKAEWNG 230
:|: :|:|:|
RESULT 21
UPK2_RHILO
ID UPK2_RHILO STANDARD; PRT; 264 AA.
AC Q98NJ1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative undecaprenol kinase 2 (EC 2.7.1.66) (Bacitracin resistance
DE protein 2).
GN UPK2 OR BACA2 OR MLR0116.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: Probably phosphorylates undecaprenol to undecaprenyl
CC phosphate. Confers resistance to bacitracin (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + undecaprenol = ADP + undecaprenyl
CC phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- MISCELLANEOUS: Bacitracin is thought to be involved in inhibition
CC of peptidoglycan synthesis by sequestering undecaprenyl
CC diphosphate reducing the pool of lipid carrier available.
CC -!- SIMILARITY: BELONGS TO THE UPK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```
-----
CC EMBL; AP002994; BAB47770.1; -.
DR HAMAP; MF_01006; -. 1.
DR InterPro; IPR003824; BacA.
DR Pfam; PF02673; BacA; 1.
KW Transferase; Kinase; Antibiotic resistance; Transmembrane;
KW Complete proteome.
FT TRANSMEM 31 50 POTENTIAL.
FT TRANSMEM 75 97 POTENTIAL.
FT TRANSMEM 107 124 POTENTIAL.
FT TRANSMEM 177 194 POTENTIAL.
FT TRANSMEM 209 231 POTENTIAL.
FT TRANSMEM 243 262 POTENTIAL.
SQ SEQUENCE 264 AA; 28109 MW; E85B086E1FDE6453 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 264;
Best Local Similarity 58.3%; Pred. No. 21;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDW 13
Db 187 PAIALAGLKELW 198

RESULT 22
ID CDX2_MOUSE STANDARD; PRT; 311 AA.
AC P43241;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein CDX-2 (Caudal-type homeobox protein 2).
GN CDX2 OR CDX-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=94253086; PubMed=7910823;
RA James R.J., Erler T., Kazenwadel J.;
RT "Structure of the murine homeobox gene cdx-2. Expression in embryonic
and adult intestinal epithelium.";
RL J. Biol. Chem. 269:15229-15237(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Small intestine;
RX MEDLINE=95021263; PubMed=7935448;
RA Suh E., Chen L., Taylor J., Traber P.G.;
RT "A homeodomain protein related to caudal regulates intestine-specific
gene transcription.";
RL Mol. Cell. Biol. 14:7340-7351(1994).
RN [3]
RP SEQUENCE OF 204-229 FROM N.A.
RX MEDLINE=91131633; PubMed=1671571;
RA James R.J., Kazenwadel J.;
RT "Homeobox gene expression in the intestinal epithelium of adult
mice.";
RL J. Biol. Chem. 266:3246-3251(1991).
RN [4]
RP PHOSPHORYLATION OF SER-60.
RX MEDLINE=21585654; PubMed=11729123;
RA Rings E.H., Boudreau F., Taylor J.K., Moffett J., Suh E.R.,
RA Traber P.G.;
RT "Phosphorylation of the serine 60 residue within the Cdx2 activation
domain mediates its transactivation capacity.";
RL Gastroenterology 121:1437-1450(2001).
CC -!- FUNCTION: Involved in the transcriptional regulation of multiple
genes expressed in the intestinal epithelium. Important in broad
range of functions from early differentiation to maintenance of
the intestinal epithelial lining of both the small and large
intestine.
CC
```

```
-----
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: INTESTINE; EXPRESSED SPECIFICALLY IN GUT
EPITHELIUM WHERE IT IS NOT RESTRICTED TO A PARTICULAR CELL
LINEAGE. ABUNDANT EXPRESSION IS SEEN IN THE PROXIMAL COLON WITH
SLIGHTLY LOWER LEVELS IN DISTAL COLON. EXPRESSION IN THE PROXIMAL
COLON IS NOT RESTRICTED EITHER TO A PARTICULAR CELL LINEAGE OR
STAGE OF DIFFERENTIATION WHILE IN THE DISTAL COLON IT IS MORE
ABUNDANT IN THE DIFFERENTIATED CELLS TOWARDS THE TOP OF THE CRYPT.
CC -!- PTM: Phosphorylation of Ser-60 mediates the transactivation
capacity.
CC -!- SIMILARITY: BELONGS TO THE CAUDAL HOMEBOX FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U00454; AAA19645.1; -.
DR EMBL; S74520; AAB32251.1; -.
DR PIR; A53808; A53808.
DR HSSP; P02835; IFTZ.
DR TRANSFAC; T02002; -.
DR MGD; MGI:88361; Cdx2.
DR GO; GO:0007389; P:pattern specification; IMP.
DR InterPro; IPR006820; Caudal_act.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH_lambrepresr.
DR Pfam; PF04731; Caudal_act; 1.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00031; HTHREPRESSR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW DNA-binding; Nuclear protein; Activator; Developmental protein; Homeobox;
FT DNA_BIND 185 244 HOMEBOX.
FT DOMAIN 46 53 POLY-ALA.
FT DOMAIN 85 92 POLY-ALA.
FT DOMAIN 247 257 POLY-GLN.
FT MOD_RES 60 60 PHOSPHORYLATION.
FT CONFLICT 69 69 Y->H (IN REF. 2).
SQ SEQUENCE 311 AA; 33476 MW; 71FFC4C263462FF3 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 311;
Best Local Similarity 56.2%; Pred. No. 25;
Matches 9; Conservative 2; Mismatches 1; Indels 4; Gaps 2;

QY 2 PTVALYG--LKQDWNG 15
Db 66 PTA--YGAPLRDWNNG 79

RESULT 23
ILVC_OCEIH
ID ILVC_OCEIH STANDARD; PRT; 344 AA.
AC Q8BN66;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
isomeroereductase) (Alpha-keto-beta-hydroxyacil reductoisomerase).
GN ILVC OR OB2621.
OS Oceanobacillus theyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
```

```

RX MEDLINE-22220767; PubMed-11235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
CC -!- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
CC = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -!- PATHWAY: Valine and isoleucine biosynthesis; second step.
CC -!- SIMILARITY: Belongs to the ketol-acid reductoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP004601; BAC14577.1; .
DR HAMAP: MF_00435; .
DR InterPro: IPR000506; ACh_isomrdctse.
DR Pfam: PF01450; ILVC; 1.
DR TIGRFAMs: TIGR00465; ilvc; 1.
KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
KW Complete proteome.
FT ACT_SITE 107 107 POTENTIAL.
SQ SEQUENCE 344 AA; 37883 MW; 467BF2E67EB8E660 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 344;
Best Local Similarity 63.6%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 ALYGLKQDWNG 15
DB 148 ALYGVHQDVTG 158

RESULT 24
DP41_BACHD STANDARD; PRT; 413 AA.
AC O9KCG7:
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase IV I (EC 2.7.7.7) (Pol IV 1).
GN DinB1 OR Bhi472.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE-20512582; PubMed-11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Poorly processive, error-prone DNA polymerase involved
CC in targeted mutagenesis. Copies undamaged DNA at stalled
CC replication forks, which arise in vivo from mismatched or
CC misaligned primer ends. These misaligned primers can be extended
CC by polIV. Exhibits no 3'-5' exonuclease (proofreading) activity.
CC May be involved in translational synthesis, in conjunction with
CC the beta clamp from polIII (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -!- COFACTOR: Binds two magnesium ions (By similarity).
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-Y FAMILY.

```

```

CC -----
CC -!- SIMILARITY: Contains 1 umuC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP001512; BAB05191.1; .
DR EMBL: H83833; H83833.1
DR HAMAP: MF_01113; 1.
DR InterPro: IPR001126; UMUC_like.
DR Pfam: PF00817; IMS; 1.
DR PROSITE: PS00173; UMUC; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW DNA-binding; Mutator protein; Magnesium; Complete proteome.
FT DOMAIN 7 188 UMUC.
FT SITE 16 16 SUBSTRATE DISCRIMINATION (BY SIMILARITY).
FT ACT_SITE 108 108 BY SIMILARITY.
FT METAL 11 11 MAGNESIUM (BY SIMILARITY).
FT METAL 107 107 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 413 AA; 47225 MW; 60F14D3356101522 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 413;
Best Local Similarity 69.2%; Pred. No. 34;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
DB 319 TVALSLWKQAWNG 331

RESULT 25
UAPA_EMENI STANDARD; PRT; 615 AA.
ID UAPA_EMENI
AC Q07307; 1994 (Rel. 30, Created)
DT 01-OCT-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Uric acid-xanthine permease (UAPA transporter).
GN UAPA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-94043131; PubMed-8226862;
RA Gorfinkiel L., Diallinas G., Scazzocchio C.;
RT "Sequence and regulation of the uapa gene encoding a uric acid-
RT xanthine permease in the fungus Aspergillus nidulans.";
RL J. Biol. Chem. 268:23376-23381(1993).
RN [2]
RP REVISIONS. G.;
RA Diallinas G.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- INDUCTION: INDUCIBLE BY 2-THIOURIC ACID, AND HIGHLY REPRESSIBLE BY
CC AMMONIUM.
CC -!- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X71807; CAA50681.1; .
DR InterPro: IPR006042; Xan_ur_permease.

```

DR InterPro: IPR006043; Xant/urac/vitC.
DR Pfam: PF00860; xan_ur_permase; 1.
DR TIGRfams: TIGR00801; ncs2; 1.
DR PROSITE: PS01116; XANTH_URACIL_PERMASE; 1.
KW Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 152 172 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 229 250 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 305 325 POTENTIAL.
FT TRANSMEM 337 356 POTENTIAL.
FT TRANSMEM 379 402 POTENTIAL.
FT TRANSMEM 468 488 POTENTIAL.
FT TRANSMEM 492 512 POTENTIAL.
FT TRANSMEM 523 543 POTENTIAL.
FT TRANSMEM 563 583 POTENTIAL.
FT CARBOHYD 10 10 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 615 AA; 65453 MW; E56A984D956897E7 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 615;
Best Local Similarity 43.5%; Pred. No. 52;
Matches 10; Conservative 2; Mismatches 5; Indels 6; Gaps 1;

OY 2 PTVALYGLK-----QDNWGVS 18
||| | | :|| | |
Db 266 PTVMLIGISLIGTGFKDWAGGSA 288

RESULT 26
PURL_ZYMMO
ID PURL_ZYMMO STANDARD; PRT; 734 AA.
AC Q9RE06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
DE synthase II).
GN PUR-O.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 31821 / ZM4 / CP4;
RA Um H.W., Kang H.S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
CC ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
CC (formamido)-N(1)-(5-phospho-D-ribosyl)acetamide + L-glutamate.
CC -!- PATHWAY: De novo purine biosynthesis; fourth step.
CC -!- SUBUNIT: Heterodimer of two subunits, purQ and purL.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF213822; AAF23789.1; -.
CC HAMAP; MF_00420; -; 1.
CC InterPro; IPR000728; AIRS_related.
CC Pfam; PF00586; AIRS; 2.
CC Pfam; PF02769; AIRS_C; 2.
KW Purine biosynthesis; Ligase; ATP-binding.
FT NP_BIND 106 117
FT SEQUENCE 734 AA; 77679 MW; B771635E0F66A166 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 734;
Best Local Similarity 58.3%; Pred. No. 62;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 PTVALYGLKQDW 13
||: || |||
Db 550 PTIGGVGLQDW 561

RESULT 27
PULA_THEMA
ID PULA_THEMA STANDARD; PRT; 843 AA.
AC Q33840;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pullulanase precursor (EC 3.2.1.41) (Alpha-dextrin endo-1,6-alpha-
DE glucosidase) (pullulan 6-glucanohydrolase).
GN PULA OR TM1845.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RA Bibel M., Brettl C., Gossler U., Kriegshaeuser G., Liebl W.;
RX MEDLINE=98115241; PubMed=9453151;
RT "Isolation and analysis of genes for amylolytic enzymes of the
RT hyperthermophilic bacterium Thermotoga maritima.";
RL FEMS Microbiol. Lett. 158:9-15(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL Nature 399:323-329(1999).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of (1->6)-alpha-D-glucosidic
CC linkages in pullulan and in amylopectin and glycogen, and the
CC alpha- and beta-limit dextrins of amylopectin and glycogen.
CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ001087; CAA04522.1; -.
CC EMBL; AE001821; AAD36907.1; -.
CC PIR; H72204; H72204.
CC TIGR; TM1845; -.
CC InterPro; IPR006047; Alpha_amyl_cat.
CC InterPro; IPR004193; Glyco_hydro_13N.
CC InterPro; IPR005323; PUD.
CC Pfam; PF00128; alpha-amylase; 1.
CC Pfam; PF02922; isoamylase.N; 1.
CC Pfam; PF03714; PUD; 1.
KW Hydrolase; Glycosidase; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 843 PULLULANASE.
FT ACT_SITE 535 535 BY SIMILARITY.
FT ACT_SITE 564 564 BY SIMILARITY.
FT ACT_SITE 652 652 BY SIMILARITY.

```

SQ SEQUENCE 843 AA; 96261 MW; C42DDE233D54FE77 CRC64;
Query Match 40.8%; Score 40; DB 1; Length 843;
Best Local Similarity 44.4%; Pred. No. 72;
Matches 8; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 1 NPTVALYGLKODWNGVSA 18
Db 556 DPTIILYGG--EPWGGWGA 571

RESULT 28
DHPL1_SCHPO STANDARD; PRT; 991 AA.
AC P40848;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein dhpl1
GN DHPL1 OR SPAC26A3.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-29.
RC STRAIN=975;
RX MEDLINE=94247347; PubMed=8190062;
RA Sugano S., Shobuike T., Takeda T., Sugino A., Ikeda H.;
RT "Molecular analysis of the dhpl1 gene of Schizosaccharomycetes pombe:
RT an essential gene that has homology to the DST2 and RAT1 genes of
RT Saccharomyces cerevisiae.";
RL Mol. Gen. Genet. 243:1-8(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks M., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welterts I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armetrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomycetes pombe.";
RL Nature 415:871-880(2002).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=21138436; PubMed=11238999;
RA Shobuike T., Tatebayashi K., Tani T., Sugano S., Ikeda H.;
RT "The dhpl1 gene, encoding a putative nuclear 5'3' exoribonuclease, is
RT required for proper chromosome segregation in fission yeast.";
RL Nucleic Acids Res. 29:1326-1333(2001).
CC -1- FUNCTION: Essential for vegetative cell growth; required for
CC proper chromosome segregation. Can bind to DNA and has 5'->3'

```

```

CC exoribonuclease activity.
CC -1- SUBUNIT: Interacts with dnl1.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the 5'-3' exonuclease family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: D17752; BAA04601.1; -
CC EMBL: Z60240; CNA93235.1; -
CC PIR: S43891; S43891.26A3.12c; -
CC GeneDB_Spombe; SPAC26A3.12c; -
CC InterPro: IPR004859; Put_53exo.
CC Pfam: PF03159; XRN_N; 1.
CC Nuclear protein; Hydrolase; Nuclease; Exonuclease.
CC KW DOMAIN 264 268 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT SEQUENCE 991 AA; 112367 MW; 83FA34D93DADAD00 CRC64;
CC -----
Query Match 40.8%; Score 40; DB 1; Length 991;
Best Local Similarity 47.1%; Pred. No. 85;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 NPTVALYGLKODWNGVS 17
Db 706 NPTIDNGKKFEWQVGA 722

RESULT 29
NKCL1_MANSE STANDARD; PRT; 1060 AA.
AC Q25479; 1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Bumetanide-sensitive sodium-(potassium)-chloride cotransporter
DE (NA-K-CL symporter).
OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Sphingioidea;
OC Sphingidae; Sphinginae; Manduca.
OX NCBI_TaxID=7130;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Malpighian tubules;
RX MEDLINE=96035837; PubMed=7550244;
RA Reagan J.D.;
RT "Molecular cloning of a putative Na(+)-K(+)-2Cl-cotransporter from
RT the Malpighian tubules of the tobacco hornworm, Manduca sexta.";
RL Insect Biochem. Mol. Biol. 25:875-880(1995).
CC -1- FUNCTION: Electrically silent transporter system. Mediates sodium
CC and chloride reabsorption. Plays a vital role in the regulation of
CC ionic balance and cell volume.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: U17344; AAA75600.1; -
CC PIR: T30823; T30823.
CC InterPro: IPR002293; AA/rel_permeasel.
CC InterPro: IPR004842; KCL_cotransp.
CC InterPro: IPR002443; NaKCL_cotransp.

```

DR PRINTS; PRO1207; NAKCLTRNSPRT.
DR TIGRFAMS; TIGR00930; 2a30; 1.
KW Transport; Ion transport; Sodium transport; Symport;
KW Potassium transport; Potassium; Transmembrane.
FT DOMAIN 1 122 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 154 174 POTENTIAL.
FT DOMAIN 175 197 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT DOMAIN 271 275 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 332 352 POTENTIAL.
FT DOMAIN 353 367 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 368 388 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT DOMAIN 453 497 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 498 518 POTENTIAL.
FT TRANSMEM 563 583 POTENTIAL.
FT DOMAIN 584 642 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 643 663 POTENTIAL.
FT TRANSMEM 882 902 POTENTIAL.
FT DOMAIN 903 1060 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1060 AA; 117392 MW; 4C39D1F0B645FFEF CRC64;

Query Match 40.8%; Score 40; DB 1; Length 1060;
Best Local Similarity 58.3%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDW 13
Db 719 PNVLLMGYKSDW 730

RESULT 30
ID Z291_HUMAN STANDARD; PRT; 1399 AA.
AC Q9BY12;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE zinc finger protein 291.
GN ZNF291
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Carim L., Sumoy L., Estivill X., Escarceller M.;
RT "Identification and characterization of ZNF291, a novel protein on
chromosome 15q24.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
DR EMBL; AF242528; AAK29205.1;
DR Genew; HGNC:13081; ZNF291.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2_1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
GN zinc-finger; Metal-binding.

FT DOMAIN 394 769 GLU-RICH.
FT ZN_FING 791 815 C2H2-TYPE.
SQ SEQUENCE 1399 AA; 158156 MW; 905C63CD908B5013 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 1399;
Best Local Similarity 43.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWGVSA 18
Db 479 SVSFCGMSMDNDVLA 494

RESULT 31
ID RL11_SPOFR STANDARD; PRT; 195 AA.
AC Q962U2;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 60S ribosomal protein L11.
GN RL11.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;
OC Noctuidae; Amphipyrinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RA Landais I., Ogliastro M., Mita K., Nohata J., Lopez-Ferber M.,
RA Duonor-Cerutti M., Fournier P., Devauchelle G.;
RT "Full-length ribosomal protein sequence from an EST library of
Spodoptera frugiperda cells (Sf9).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY ASSOCIATES WITH 5S RNA.
CC -!- SIMILARITY: BELONGS TO THE L5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
DR EMBL; AF400182; AAK92154.1;
DR InterPro; IPR002132; Ribosomal_L5.
DR InterPro; IPR003236; Ribosomal_L5_mit.
DR Pfam; PF00281; Ribosomal_L5; 1.
DR Pfam; PF00673; Ribosomal_L5_C; 1.
DR PROSITE; PS00358; RIBOSOMAL_L5; 1.
KW Ribosomal protein; rRNA-binding.
SQ SEQUENCE 195 AA; 22346 MW; A34A304552E1BF5B5 CRC64;

Query Match 40.3%; Score 39.5; DB 1; Length 195;
Best Local Similarity 25.0%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 3; Indels 17; Gaps 1;

QY 1 NPTVALYGL-----KQDWNG 15
Db 135 DPSIGYGLDFVVLGPTRIQCTTKTQDQWG 166

RESULT 32
ID ADB2_YEAST STANDARD; PRT; 726 AA.
AC P36000;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable beta-adaptin (Clathrin assembly protein large beta chain)
DE (Clathrin assembly protein complex 2 beta large chain).
GN APL2 OR YKL135C.

```
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95340689; PubMed=7615679;
RA Rad M.R., Phan H.L., Kirchhuth L., Tan P.K., Kirchhausen T.,
RA Hollenberg C.P., Payne G.S.;
RA "Saccharomyces cerevisiae APl2p, a homologue of the mammalian
RT clathrin AP beta subunit, plays a role in clathrin-dependent Golgi
RT functions.";
RL J. Cell Sci. 108:1605-1615(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Rad M.R., Xu G., Kirchhuth L., Fritz C., Keuchel H., Hollenberg C.P.;
RA Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RL
CC -!- FUNCTION: ADAPTINS ARE COMPONENTS OF THE ADAPTOR COMPLEXES
CC WHICH LINK CLATHRIN TO RECEPTORS IN COATED VESICLES. CLATHRIN-
CC ASSOCIATED PROTEIN COMPLEXES ARE BELIEVED TO INTERACT WITH THE
CC CYTOPLASMIC TAILS OF MEMBRANE PROTEINS, LEADING TO THEIR SELECTION
CC AND CONCENTRATION. BETA ADAPTIN IS A SUBUNIT OF THE PLASMA
CC MEMBRANE ADAPTOR.
CC -!- SUBUNIT: ASSEMBLY PROTEIN COMPLEX 2 (AP-2) IS A HETEROTETRAMER
CC COMPOSED OF TWO LARGE CHAINS (ALPHA AND BETA), A MEDIUM CHAIN
CC (AP54) AND A SMALL CHAIN (AP17).
CC -!- SUBCELLULAR LOCATION: COMPONENT OF THE COAT SURROUNDING THE
CC CYTOPLASMIC FACE OF COATED VESICLES IN THE PLASMA MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNITS
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z30212; CAAB2931.1; -.
CC EMBL; Z28135; CAAB1977.1; -.
CC PIR; S37964; S37964.
CC SGD; S0001618; APL2.
CC GO; GO:0030121; C:AP-1 adaptor complex; IPI.
CC GO; GO:0030276; F:clathrin binding activity; IPI.
CC GO; GO:0016192; P:vesicle-mediated transport; IPI.
CC InterPro; IPR002553; Adaptin_N.
CC Pfam; PF01602; Adaptin_N; 1.
CC Coated pits.
CC SEQUENCE 726 AA; 81868 MW; 89893945D4EC6C4A CRC64;
AC Q49116
IN YMEREX STANDARD; PRT; 228 AA.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 23.6 kDa protein in mea3 region precursor (ORFB).
OS Methylobacterium extorquens.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Methylorhizobactaceae; Methylobacterium.
OC NCBI_TaxID=408;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=AM1 / NCIMB 9133;
RA MEDLINE=97022083; PubMed=8868443;
RA Smith L.M., Meijer W.G., Dijkhuizen L., Goodwin P.M.;
RA "A protein having similarity with methylmalonyl-CoA mutase is
RT required for the assimilation of methanol and ethanol by
RT Methylobacterium extorquens AM1.";
RL Microbiology 142:675-684(1996).
RN
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28335; AAC44088.1; -.
CC Hypothetical protein: Signal.
CC SIGNAL 1 23
CC CHAIN 24 228
CC HYPOTHETICAL 25.6 kDa PROTEIN IN MEAA
CC SEQUENCE 228 AA; 25577 MW; 4EEF22FEB061A9B4 CRC64;
DR
KW Hypothetical protein: Signal.
FT SIGNAL 1 23
FT CHAIN 24 228
FT
SQ SEQUENCE 228 AA; 25577 MW; 4EEF22FEB061A9B4 CRC64;
Query Match 39.8%; Score 39; DB 1; Length 228;
Best Local Similarity 70.0%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 9 LKQDNGVSA 18
DB 149 LPKDNGVSA 158
ID PYRF_LACLC STANDARD; PRT; 237 AA.
AC P50924;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase)
DE (OMPDecase) (OMPDecase).
DE GN PYRF.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OC NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=MG1363;
RA MEDLINE=96326348; PubMed=8759867;
RA Andersen P.S., Martinussen J., Hammer K.;
RA "Sequence analysis and identification of the pyrKdbf operon from
RT Lactococcus lactis including a novel gene, pyrK, involved in
RT pyrimidine biosynthesis.";
RL J. Bacteriol. 178:5005-5012(1996).
CC -!- CATALYTIC ACTIVITY: Orotidine 5'-phosphate = UMP + CO(2).
CC -!- PATHWAY: Pyrimidine biosynthesis; sixth (last) step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE OMP DECARBOXYLASE FAMILY. SUBFAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74207; CAAS2281.1; -.
CC HSSP; P25971; IDBT.
CC HAMAP; MF_01200; -.
CC InterPro; IPR001754; OMPDecase.
CC Pfam; PF00215; OMPDecase; 1.
DR
```

DR PROSITE; PS00156; OMPDECASE; 1.
 KW Lyase; Decarboxylase; Pyrimidine biosynthesis.
 FT ACT_SITE 63 63 PROTON DONOR (BY SIMILARITY).
 SQ SEQUENCE 237 AA; 26169 MW; F799EE4BCD7A1A54 CRC64;

Query Match 39.8%; Score 39; DB 1; Length 237;
 Best Local Similarity 35.7%; Pred. No. 28;
 Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWN 14
 :| | | | | | | | | |
 Db 223 DPVSAYHAIKEWN 236

RESULT 35
 OMK2_VIBPA
 ID OMK2_VIBPA STANDARD; PRT; 263 AA.
 AC P51002;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Outer membrane protein ompK precursor.
 GN OMPK.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RIMD 2210001;
 RX MEDLINE=96102868; PubMed=8586275;
 RA Inoue T., Matsuzaki S., Tanaka S.;
 RT "Cloning and sequence analysis of Vibrio parahaemolyticus ompK gene
 encoding a 26-kDa outer membrane protein, OmpK, that serves as
 receptor for a broad-host-range vibriophage, KVP40.";
 RL FEMS Microbiol. Lett. 134:245-249(1995).
 CC -!- FUNCTION: Serves as receptor for a broad-host-range vibriophage,
 KVP40.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: SOME, TO E.COLI TSX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D61392; BAA09613.1; -
 DR InterPro; IPR003055; Channel_Tsx.
 DR Pfam; PF03502; Channel_Tsx; 1.
 KW Signal; Outer membrane; Phage recognition.
 FT SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 263 OUTER MEMBRANE PROTEIN OMPK.
 SQ SEQUENCE 263 AA; 29467 MW; 1CAB3EA7BADB447B CRC64;

Query Match 39.8%; Score 39; DB 1; Length 263;
 Best Local Similarity 57.1%; Pred. No. 31;
 Matches 8; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

QY 6 LYGL----KQDWNG 15
 ||| | | | | |
 Db 158 LYGYDSNKKDWNG 171

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 27.1642 Seconds
(without alignments)
189.995 Million cell updates/sec

Title: US-09-765-739a-1

Perfect score: 110

Sequence: 1 KSTVGVFGLKHDWDGSPILK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_23:*

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.prodent:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.rvirus:*
- 16: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	ID	Description
1	110	100.0	278	Q9FA72	Q9f472 ehrlichia c
2	110	100.0	278	Q9RAA8	Q9r8a8 ehrlichia c
3	110	100.0	278	Q9RAA7	Q9r8a7 ehrlichia c
4	110	100.0	278	Q9R3J3	Q9r3j3 ehrlichia c
5	110	100.0	278	Q9RAA6	Q9r8a6 ehrlichia c
6	110	100.0	278	Q9RAA9	Q9r8a9 ehrlichia c
7	110	100.0	278	Q9RAA5	Q9r8a5 ehrlichia c
8	110	100.0	307	Q9ZGJ1	Q9zgj1 ehrlichia c
9	79	71.8	280	Q9ADV3	Q9adv3 ehrlichia c
10	79	71.8	280	Q9F473	Q9f473 ehrlichia c
11	76	69.1	280	Q52107	Q52107 ehrlichia c
12	73	66.4	246	Q9RH35	Q9rh35 ehrlichia c
13	73	66.4	275	Q93DD4	Q93dd4 ehrlichia c
14	73	66.4	276	Q93DD1	Q93dd1 ehrlichia c
15	73	66.4	276	Q85817	Q85817 ehrlichia c
16	73	66.4	276	Q8GGU0	Q8gguo ehrlichia c

Q93dd3	ehrllichia c	280	2	Q93DD3
Q9zgm9	ehrllichia c	280	2	Q9ZGM9
Q85816	ehrllichia c	280	2	O85816
Q93dd2	ehrllichia c	281	2	Q93DD2
Q9aci9	ehrllichia c	281	2	Q9ACI9
Q9zgj2	ehrllichia c	288	2	Q9ZGJ2
Q9aeu3	cowdria rum	290	2	Q9AEU3
Q9af99	cowdria rum	265	2	Q9AF99
Q9af98	cowdria rum	270	2	Q9AF98
Q93e59	cowdria rum	275	2	Q93E59
Q93e60	cowdria rum	276	2	Q93E60
Q93e53	cowdria rum	276	2	Q93E53
Q93e65	cowdria rum	277	2	Q93E65
Q93e52	cowdria rum	278	2	Q93E52
Q93e57	cowdria rum	278	2	Q93E57
Q93e56	cowdria rum	278	2	Q93E56
Q46329	cowdria rum	287	2	Q46329
Q9r425	cowdria rum	287	2	Q9R425
Q46331	cowdria rum	287	2	Q46331
Q46324	cowdria rum	290	2	Q46324
Q93e64	cowdria rum	290	2	Q93E64
Q46333	cowdria rum	290	2	Q46333
Q46332	cowdria rum	290	2	Q46332
Q46330	cowdria rum	290	2	Q46330
Q9amf6	ehrllichia s	272	2	Q9AMF6
Q93e54	cowdria rum	272	2	Q93E54
Q9afal	cowdria rum	284	2	Q9AFAL
Q46327	cowdria rum	284	2	Q46327
Q52105	ehrllichia c	286	2	Q52105
Q8ggq3	ehrllichia c	287	2	Q8GGQ3
Q8ggq2	ehrllichia c	291	2	Q8GGQ2
Q8g921	ehrllichia c	291	2	Q8G921
Q9f475	ehrllichia c	276	2	Q9F475
Q93e62	cowdria rum	269	2	Q93E62
Q9s6h1	cowdria rum	281	2	Q9S6H1
Q46328	cowdria rum	281	2	Q46328
Q9s6h0	cowdria rum	281	2	Q9S6H0
Q8g948	ehrllichia c	276	2	Q8G948
Q8ggul	ehrllichia c	277	2	Q8GGUL
Q8g8w7	ehrllichia c	277	2	Q8G8W7
Q93e55	cowdria rum	280	2	Q93E55
Q93e58	cowdria rum	280	2	Q93E58
Q9afao	cowdria rum	271	2	Q9AFAO
Q52106	ehrllichia c	278	2	Q52106
Q8g8q5	ehrllichia c	278	2	Q8G8Q5
Q8g8p3	ehrllichia c	291	2	Q8G8P3
Q52104	ehrllichia c	280	2	Q52104
Q8ggq5	ehrllichia c	280	2	Q8GGQ5
Q8g9j3	ehrllichia c	280	2	Q8G9J3
Q9fu09	phaseolus v	904	10	Q9FU09
Q85360	ehrllichia c	133	2	O85360
Q93e61	cowdria rum	268	2	Q93E61
Q93e63	cowdria rum	268	2	Q93E63
Q9f474	ehrllichia c	283	2	Q9F474
Q8t882	arabella ir	397	5	Q8T882
Q8ud44	agrobacteri	385	16	Q8UD44
Q8bmn4	mus musculu	681	11	Q8BMN4
Q9f221	helicobacte	1165	2	Q9F221
Q915x9	helicobacte	1179	2	Q915X9
Q8kzh3	helicobacte	1180	2	Q8KZH3
Q86064	helicobacte	1183	2	O86064
Q8kzh9	helicobacte	1216	2	Q8KZH9
Q9f223	helicobacte	1247	2	Q9F223
Q07910	helicobacte	1247	2	Q07910
Q9rf15	helicobacte	1247	2	Q9RF15
O16875	caenorhabdi	231	5	O16875
Q9fiq9	arabidopsis	305	10	Q9FIQ9
Q9uy88	pyrococcus	103	17	Q9UY88
Q57197	staphylococ	132	2	Q57197
O05155	staphylococ	132	16	O05155
O52167	salmonella	144	2	O52167
Q8nbc5	homo sapien	199	4	Q8NBC5
Q8nia6	homo sapien	199	4	Q8NIA6

90 Q9sa03 arabidopsals
 91 Q38544 bacterioph
 92 Q8rbz6 thermoanaer
 93 Q9t122 bacterioph
 94 Q8eg37 shewanella
 95 Q96kr5 homo sapien
 96 Q96kr4 homo sapien
 97 Q9t123 bacterioph
 98 Q9a500 caulobacter
 99 Q8bj70 mus musculu
 100 Q8x4h5 escherichia

ALIGNMENTS

RESULT 1

Q9F472 ID Q9F472 PRELIMINARY; PRT; 278 AA.
 AC Q9F472; 2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE P28-7.
 GN P28-7.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Jake;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 antigen."
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Jake;
 RX MEDLINE=20432107; PubMed=10974556;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "A conserved, transcriptionally active p28 multigene locus of
 Ehrlichia canis."
 RL Gene 254:245-252(2000).
 DR EMBL; AF082744; AAC64550.2; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT Non_TER 278 278
 SQ SEQUENCE 278 AA; 30455 MW; 2411CAAB4C56CA74 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;
 Best Local Similarity 100.0%; Pred. No. 2.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTVGVLKHDWDGSPILK 20
 Db 60 KSTVGVLKHDWDGSPILK 79

RESULT 2

Q9R8A8 ID Q9R8A8 PRELIMINARY; PRT; 278 AA.
 AC Q9R8A8; 2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.

RP SEQUENCE FROM N.A.
 RC STRAIN=Oklahoma;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 antigen."
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL; AF082744; AAC64552.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT Non_TER 278 278
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;
 Best Local Similarity 100.0%; Pred. No. 2.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTVGVLKHDWDGSPILK 20
 Db 60 KSTVGVLKHDWDGSPILK 79

RESULT 3

Q9R8A7 ID Q9R8A7 PRELIMINARY; PRT; 278 AA.
 AC Q9R8A7; 2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Jemon;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 antigen."
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL; AF082747; AAC64553.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT Non_TER 278 278
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;
 Best Local Similarity 100.0%; Pred. No. 2.8e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSTVGVLKHDWDGSPILK 20
 Db 60 KSTVGVLKHDWDGSPILK 79

RESULT 4

Q9R3J3 ID Q9R3J3 PRELIMINARY; PRT; 278 AA.
 AC Q9R3J3; 2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RC STRAIN=Florida;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082750; AAC64556.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTGVGFLKHDWDGSPILK 20
Db 60 KSTGVGFLKHDWDGSPILK 79

RESULT 5
Q9R8A6
ID Q9R8A6 PRELIMINARY; PRT; 278 AA.
AC Q9R8A6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DJ;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082748; AAC64554.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTGVGFLKHDWDGSPILK 20
Db 60 KSTGVGFLKHDWDGSPILK 79

RESULT 6
Q9R8A9
ID Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Louisiana;
```

```
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082745; AAC64551.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTGVGFLKHDWDGSPILK 20
Db 60 KSTGVGFLKHDWDGSPILK 79

RESULT 7
Q9R8A5
ID Q9R8A5 PRELIMINARY; PRT; 278 AA.
AC Q9R8A5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fuzzy;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL; AF082749; AAC64555.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 100.0%; Score 110; DB 2; Length 278;
Best Local Similarity 100.0%; Pred. No. 2.8e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTGVGFLKHDWDGSPILK 20
Db 60 KSTGVGFLKHDWDGSPILK 79

RESULT 8
Q9ZGJ1
ID Q9ZGJ1 PRELIMINARY; PRT; 307 AA.
AC Q9ZGJ1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major outer membrane protein p30-1.
GN P30-1.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
```



```
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allenan A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: U72291; AAC02940.1; -.
DR EMBL: AF479833; AAC012931.1; -.
DR EMBL: AF479834; AAC012937.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;

Query Match 69.1%; Score 76; DB 2; Length 280;
Best Local Similarity 73.7%; Pred. No. 0.0069;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPILK 20
Db :| ||||| ||||| |
61 TTGVFGLKQDWDGSAI 79

RESULT 12
Q9RH35 PRELIMINARY; PRT; 246 AA.
AC Q9RH35;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (Fragment).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St.Vincent.
RX MEDLINE=99175287; PubMed=10074538;
RA Xu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 66.4%; Score 73; DB 2; Length 246;
Best Local Similarity 76.5%; Pred. No. 0.0018;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPIL 18
Db :| ||||| ||||| |
30 TTGVFGLKQDWDGSAI 46

RESULT 15
Q85817 PRELIMINARY; PRT; 276 AA.
ID Q85817
AC Q85817;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
```

```
RESULT 13
Q93DD4 PRELIMINARY; PRT; 275 AA.
ID Q93DD4;
AC Q93DD4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V2;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393389; AAL12919.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 29974 MW; 2ECCF2F988B2E9D9 CRC64;

Query Match 66.4%; Score 73; DB 2; Length 275;
Best Local Similarity 76.5%; Pred. No. 0.002;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPIL 18
Db :| ||||| ||||| |
60 TTGVFGLKQDWDGSAI 76

RESULT 14
Q93DD1 PRELIMINARY; PRT; 276 AA.
ID Q93DD1;
AC Q93DD1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393393; AAL12923.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30028 MW; 2D7143AFCBF2EBE CRC64;

Query Match 66.4%; Score 73; DB 2; Length 276;
Best Local Similarity 76.5%; Pred. No. 0.002;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPIL 18
Db :| ||||| ||||| |
60 TTGVFGLKQDWDGSAI 76

RESULT 15
Q85817 PRELIMINARY; PRT; 276 AA.
ID Q85817
AC Q85817;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
```



```

DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Query Match 66.4%; Score 73; DB 2; Length 280;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
   :| ||||| ||||| |
Db 60 TTAGVFGKQDWDGSAI 76

RESULT 19
O85816 PRELIMINARY; PRT; 280 AA.
ID O85816;
AC O85816;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28kDa outer membrane protein gene
19).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RX PubMed=12496165;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=v8, and V4;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Outer Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077732; AAC31545.1; -
DR EMBL; AF393394; AAL12924.1; -
DR EMBL; AF393390; AAL12920.1; -
DR EMBL; AF479835; AAO12964.1; -
DR EMBL; AF479840; AAO12970.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 66.4%; Score 73; DB 2; Length 280;
Best Local Similarity 76.5%; Pred. No. 0.0021;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
   :| ||||| ||||| |
Db 60 TTAGVFGKQDWDGSAI 76

us-09-765-739a-1.rspt

```

```

RESULT 20
Q93DD2 PRELIMINARY; PRT; 281 AA.
ID Q93DD2;
AC Q93DD2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V6;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393392; AAL12922.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30388 MW; B54DBB55AC839A9A CRC64;

Query Match 65.5%; Score 72; DB 2; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.003;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
   :| ||||| ||||| |
Db 60 TTGVFGKQNDGSAI 76

RESULT 21
Q9AC19 PRELIMINARY; PRT; 281 AA.
ID Q9AC19;
AC Q9AC19;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Major outer membrane protein P28 (28 kDa outer membrane protein).
GN P28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V1;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Lithonia, Arkansas, and Osciola;

```

EX PubMed-12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allenan A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichiae";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL: U72291; AAK28673.1;
DR EMBL: AF393388; AAL12018.1;
DR EMBL: AF117396; AAK77031.1;
DR EMBL: AF479833; AAK12932.1;
DR EMBL: AF479834; AAK12938.1;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 65.5%; Score 72; DB 2; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.003;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
Db 60 TTGVVFLKQNDGSAI 76
:|||||:|||||

RESULT 22
Q9ZGJ2 PRELIMINARY; PRT; 288 AA.
AC Q9ZGJ2
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 30-kDa major outer membrane protein (P28-8).
GN P30 OR P28-8.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OKLAHOMA;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
immunodominant 30-kilodalton major outer membrane proteins of
Ehrlichia canis, and application of the recombinant protein for
serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
Ehrlichia canis.";
RL Gene 254:245-252(2000).

DR EMBL: AF078553; AAC68667.1; -;
DR EMBL: AF082744; AAG14362.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BF5E CRC64;

Query Match 63.6%; Score 70; DB 2; Length 288;
Best Local Similarity 70.6%; Pred. No. 0.0063; 3; Indels 0; Gaps 0;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
Db 61 TTGVVFLKQNDGAT1 77
:|||||:|||||

RESULT 23
Q9AEU3 PRELIMINARY; PRT; 290 AA.
AC Q9AEU3
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Major antigenic protein MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vosloo;
RA Bensaïd A., Allsopp M.T., Maillard J.-C., Chantal I., Allsopp B.;
RT "Cowdria ruminantium major antigenic protein (map1) gene variants are
not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY028378; AAK18727.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 290 AA; 31462 MW; FBACF99309E10AFF CRC64;

Query Match 57.3%; Score 63; DB 2; Length 290;
Best Local Similarity 80.0%; Pred. No. 0.082;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVEGLKHDWDG 15
Db 61 KSTQTVFGLKHDWDG 75
||| |||||

RESULT 24
Q9AF99 PRELIMINARY; PRT; 265 AA.
AC Q9AF99
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Major antigenic protein 1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Umpala;
RA Bensaïd A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF355202; AAK27218.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 265 265
FT SEQUENCE 265 AA; 28388 MW; 1EA6DB25CCBF9634 CRC64;

Query Match 54.5%; Score 60; DB 2; Length 265;
Best Local Similarity 73.3%; Pred. No. 0.22;
Matches 11; Conservative 1; Mismatches 0; Indels 3; Gaps 0;

QY 1 KSTVGVLKHDWDG 15
I:I |||||
DB 43 KNTQTVGLKDWG 57

RESULT 25

Q9AF98 ID Q9AF98 PRELIMINARY; PRT; 270 AA.
AC Q9AF98;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Major antigenic protein 1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cameroun;
RA Bensaïd A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein 1 (map1) gene variants
are not geographically constrained."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF355203; AAK27219.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 270 270
SQ SEQUENCE 270 AA; 29380 MW; 99BNAFE9E48895A6 CRC64;

Query Match 54.5%; Score 60; DB 2; Length 270;
Best Local Similarity 73.3%; Pred. No. 0.23;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVLKHDWDG 15
I:I |||||
DB 47 KNTQTVGLKDWG 61

RESULT 26

Q93E59 ID Q93E59 PRELIMINARY; PRT; 275 AA.
AC Q93E59;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mar8777;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
Are Not Geographically Constrained and Show No Evidence of Having
Evolved under Positive Selection Pressure."
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368008; AAK98148.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 275 275
SQ SEQUENCE 275 AA; 29722 MW; 71170E2A159B28DB CRC64;

Query Match 54.5%; Score 60; DB 2; Length 275;
Best Local Similarity 73.3%; Pred. No. 0.23;
Matches 11; Conservative 1; Mismatches 0; Indels 3; Gaps 0;

QY 1 KSTVGVLKHDWDG 15
I:I |||||
DB 55 KNTQTVGLKDWG 69

RESULT 27

Q93E60 ID Q93E60 PRELIMINARY; PRT; 276 AA.
AC Q93E60;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mali;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
Are Not Geographically Constrained and Show No Evidence of Having
Evolved under Positive Selection Pressure."
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368007; AAK98147.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 276 276
SQ SEQUENCE 276 AA; 29856 MW; F685387E1602A50E CRC64;

Query Match 54.5%; Score 60; DB 2; Length 276;
Best Local Similarity 73.3%; Pred. No. 0.23;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVLKHDWDG 15
I:I |||||
DB 55 KNTQTVGLKDWG 69

RESULT 28

Q93E53 ID Q93E53 PRELIMINARY; PRT; 276 AA.
AC Q93E53;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SanKat;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
Are Not Geographically Constrained and Show No Evidence of Having
Evolved under Positive Selection Pressure."
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368014; AAK98154.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.

```

FT  NON_TER      1
FT  NON_TER      276
SQ  SEQUENCE      276 AA; 29870 MW; E415387E1602A50E CRC64;

Query Match      54.5%; Score 60; DB 2; Length 276;
Best Local Similarity 73.3%; Pred. No. 0.23;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY  1 KSTVGVEGLKHDWDG 15
Db  55 KNTQTVEGLKHDWDG 69

RESULT 29
ID  Q93E65          PRELIMINARY; PRT; 277 AA.
AC  Q93E65;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DE  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE  01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE  Major antigenic protein MAP1 (Fragment).
OS  Cowdria ruminantium.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Anaplasmataceae; Ehrlichia.
OX  NCBI_TaxID=779;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Blaukrans;
RX  MEDLINE=21539003; PubMed=11682561;
RA  Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA  Haydon D.T., van Heerden H., Allsopp B.A.;
RT  "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT  Are Not Geographically Constrained and Show No Evidence of Having
RT  Evolved under Positive Selection Pressure.";
RL  J. Clin. Microbiol. 39:4200-4203(2001).
DR  EMBL; AF368001; AAK98142.1;
DR  InterPro; IPR002566; Surface_Ag_msp4.
DR  Pfam; PF01617; Surface_Ag_2; 1.
FT  NON_TER      277
FT  NON_TER      277
SQ  SEQUENCE      277 AA; 29916 MW; 167B95113194237A CRC64;

Query Match      54.5%; Score 60; DB 2; Length 277;
Best Local Similarity 73.3%; Pred. No. 0.23;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY  1 KSTVGVEGLKHDWDG 15
Db  54 KNTQTVEGLKHDWDG 68

RESULT 30
ID  Q93E52          PRELIMINARY; PRT; 278 AA.
AC  Q93E52;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DE  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE  01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE  Major antigenic protein MAP1 (Fragment).
OS  Cowdria ruminantium.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Anaplasmataceae; Ehrlichia.
OX  NCBI_TaxID=779;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=South East Botswana;
RX  MEDLINE=21539003; PubMed=11682561;
RA  Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA  Haydon D.T., van Heerden H., Allsopp B.A.;
RT  "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT  Are Not Geographically Constrained and Show No Evidence of Having
RT  Evolved under Positive Selection Pressure.";
RL  J. Clin. Microbiol. 39:4200-4203(2001).

```

```

DR  EMBL; AF368015; AAK98155.1;
DR  InterPro; IPR002566; Surface_Ag_msp4.
DR  Pfam; PF01617; Surface_Ag_2; 1.
FT  NON_TER      278
FT  NON_TER      278
SQ  SEQUENCE      278 AA; 30156 MW; 0355DE6B85842386 CRC64;

Query Match      54.5%; Score 60; DB 2; Length 278;
Best Local Similarity 73.3%; Pred. No. 0.23;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY  1 KSTVGVEGLKHDWDG 15
Db  55 KNTQTVEGLKHDWDG 69

RESULT 31
ID  Q93E57          PRELIMINARY; PRT; 278 AA.
AC  Q93E57;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DE  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE  01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE  Major antigenic protein MAP1 (Fragment).
OS  Cowdria ruminantium.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Anaplasmataceae; Ehrlichia.
OX  NCBI_TaxID=779;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Morgenswad;
RX  MEDLINE=21539003; PubMed=11682561;
RA  Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA  Haydon D.T., van Heerden H., Allsopp B.A.;
RT  "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT  Are Not Geographically Constrained and Show No Evidence of Having
RT  Evolved under Positive Selection Pressure.";
RL  J. Clin. Microbiol. 39:4200-4203(2001).
DR  EMBL; AF368010; AAK98150.1;
DR  InterPro; IPR002566; Surface_Ag_msp4.
DR  Pfam; PF01617; Surface_Ag_2; 1.
FT  NON_TER      278
FT  NON_TER      278
SQ  SEQUENCE      278 AA; 30048 MW; 6C74669F156D4980 CRC64;

Query Match      54.5%; Score 60; DB 2; Length 278;
Best Local Similarity 73.3%; Pred. No. 0.23;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY  1 KSTVGVEGLKHDWDG 15
Db  55 KNTQTVEGLKHDWDG 69

RESULT 32
ID  Q93E56          PRELIMINARY; PRT; 278 AA.
AC  Q93E56;
DT  01-DEC-2001 (TrEMBLrel. 19, Created)
DE  01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE  01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE  Major antigenic protein MAP1 (Fragment).
OS  Cowdria ruminantium.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Anaplasmataceae; Ehrlichia.
OX  NCBI_TaxID=779;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Noneto;
RX  MEDLINE=21539003; PubMed=11682561;
RA  Allsopp M.T.E.P., Dorfling C.M., Maillard J.C., Bensaïd A.,
RA  Haydon D.T., van Heerden H., Allsopp B.A.;
RT  "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants

```

```
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J. Clin. Microbiol. 39:4200-4203(2001).
DR EMBL; AF368011; AAK98151.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30153 MW; 368C3C505C0A2DC6 CRC64;

Query Match 54.5%; Score 60; DB 2; Length 278;
Best Local Similarity 73.3%; Pred. No. 0.24;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVFGLKHDWDG 15
   I: I I I I I I I I I
Db 55 KNTQTVFGLKDWGD 69

RESULT 33
Q46329 PRELIMINARY; PRT; 287 AA.
AC Q46329;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major antigenic protein 1.
GN MAPI.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Crystal Springs;
RX MEDLINE=96400830; PubMed=8807206;
RA Reddy G.R., Sulsona C.R., Harrison R.H., Mahan S.M., Burrridge M.J.,
RA Barbet A.F.;
RT "Sequence heterogeneity of the major antigenic protein 1 genes from
RT Cowdria ruminantium isolates from different geographical areas.";
RL Clin. Diagn. Lab. Immunol. 3:417-422(1996).
DR EMBL; U50831; AAC44142.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 287 AA; 30949 MW; 4509731A782F4666 CRC64;

Query Match 54.5%; Score 60; DB 2; Length 287;
Best Local Similarity 73.3%; Pred. No. 0.24;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVFGLKHDWDG 15
   I: I I I I I I I I I
Db 61 KNTQTVFGLKDWGD 75

RESULT 34
Q9R425 PRELIMINARY; PRT; 287 AA.
AC Q9R425;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Major antigenic protein 1.
GN MAPI.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HIGHWAY, and CRYSTAL SPRINGS;
RX MEDLINE=99216274; PubMed=10198207;
RA Sulsona C.R., Mahan S.M., Barbet A.F.;
```

```
RT "The map1 Gene of Cowdria ruminantium Is a Member of a Multigene
RT Family Containing Both Conserved and Variable Genes.";
RL Biochem. Biophys. Res. Commun. 257:300-305(1999).
DR EMBL; AF125276; AAD26348.1; -.
DR EMBL; AF125275; AAD26346.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 287 AA; 31104 MW; 15FBE53BC8215F69 CRC64;

Query Match 54.5%; Score 60; DB 2; Length 287;
Best Local Similarity 73.3%; Pred. No. 0.24;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVFGLKHDWDG 15
   I: I I I I I I I I I
Db 61 KNTQTVFGLKDWGD 75

RESULT 35
Q46331 PRELIMINARY; PRT; 287 AA.
AC Q46331;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Major antigenic protein 1.
GN MAPI.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Highway;
RX MEDLINE=96400830; PubMed=8807206;
RA Reddy G.R., Sulsona C.R., Harrison R.H., Mahan S.M., Burrridge M.J.,
RA Barbet A.F.;
RT "Sequence heterogeneity of the major antigenic protein 1 genes from
RT Cowdria ruminantium isolates from different geographical areas.";
RL Clin. Diagn. Lab. Immunol. 3:417-422(1996).
DR EMBL; U50833; AAC44144.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 287 AA; 31120 MW; A0E321EE49FD5F74 CRC64;

Query Match 54.5%; Score 60; DB 2; Length 287;
Best Local Similarity 73.3%; Pred. No. 0.24;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVFGLKHDWDG 15
   I: I I I I I I I I I
Db 61 KNTQTVFGLKDWGD 75

Search completed: October 6, 2003, 07:52:28
Job time : 34.1642 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:03 ; Search time 35.3731 Seconds
(without alignments)
89.744 Million cell updates/sec

Title: US-09-765-739A-2

Perfect score: 110

Sequence: 1 NTTCGVFLKQDWDGATIKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_13Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	110	100.0	20	ABG30744 Ehrlichia canis pe
2	110	100.0	288	RAY08959 E. canis p30 prote
3	110	100.0	288	ABG77950 Ehrlichia canis ou
4	96	87.3	19	ABG30749 Ehrlichia chaffeen
5	96	87.3	280	RAY06948 E. chafeensis OMP-
6	96	87.3	280	ABG77940 Ehrlichia chaffeen
7	96	87.3	280	23 AAU96110 Ehrlichia chafeens
8	90	81.8	280	19 AAWS1094 Ehrlichia chaffeen
9	90	81.8	280	21 AAB36188 Ehrlichia chaffeen

90	81.8	280	22	AAU04198	Variable surface a
90	81.8	280	23	AAU73417	Ehrlichia chaffeen
82	74.5	20	23	ABG30745	Ehrlichia chaffeen
13	74.5	256	20	RAY06942	E. chafeensis p28
14	74.5	256	23	ABG77966	Protein encoded by
15	74.5	276	19	AAWS1095	Ehrlichia chaffeen
16	74.5	276	21	AAB36189	Ehrlichia chaffeen
17	74.5	276	22	AAU04199	Variable surface a
18	74.5	280	19	AAWS1089	Ehrlichia chaffeen
19	74.5	280	21	AAB36183	Ehrlichia chaffeen
20	74.5	280	22	AAU04193	Major antigenic pr
21	74.5	281	20	AAU06943	E. chafeensis OMP-
22	74.5	281	23	ABG77935	Ehrlichia chaffeen
23	74.5	281	23	AAU96105	Ehrlichia chafeens
24	74.5	281	23	AAU73418	Ehrlichia chaffeen
25	70.0	280	20	AAU06962	E. canis p30-2 pro
26	70.0	280	21	AAU71479	Ehrlichia canis im
27	70.0	280	23	ABG77953	Ehrlichia canis ou
28	70.0	280	23	AAU96102	Ehrlichia canis p2
29	69.1	276	20	AAU06964	E. canis p30-4 pro
30	69.1	276	23	ABG77955	Ehrlichia canis ou
31	69.1	276	23	AAU96117	Ehrlichia canis p2
32	64.5	19	23	ABG30747	Ehrlichia chaffeen
33	64.5	286	19	AAWS1092	Ehrlichia chaffeen
34	64.5	286	20	AAU06946	E. chafeensis OMP-
35	64.5	286	21	AAB36186	Ehrlichia chaffeen
36	64.5	286	22	AAU04196	Variable surface a
37	64.5	286	23	ABG77938	Ehrlichia chaffeen
38	64.5	286	23	AAU96108	Ehrlichia chafeens
39	64.5	286	23	AAU73415	Ehrlichia chaffeen
40	63.6	20	23	ABG30743	Ehrlichia canis pe
41	63.6	278	21	AAU71477	Ehrlichia canis im
42	63.6	278	23	AAU96100	Ehrlichia canis p2
43	63.6	307	20	AAU06961	E. canis p30-1 pro
44	63.6	307	23	ABG77952	Ehrlichia canis ou
45	54.1	287	19	AAWS1088	Cowdria ruminantiu
46	54.1	287	21	AAB36182	Cowdria ruminantiu
47	54.1	287	22	AAU04192	Major antigenic pr
48	53.6	18	23	ABG30746	Ehrlichia chaffeen
49	53.6	18	23	ABG30748	Ehrlichia chaffeen
50	53.6	278	19	AAWS1093	Ehrlichia chaffeen
51	53.6	278	20	AAU06947	Ehrlichia chaffeen
52	53.6	278	21	AAB36187	Ehrlichia chaffeen
53	53.6	278	22	AAU04197	Variable surface a
54	53.6	278	23	AAE31090	Ehrlichia ruminant
55	53.6	278	23	ABG77939	Ehrlichia chaffeen
56	53.6	278	23	AAU96109	Ehrlichia chafeens
57	53.6	278	23	AAU73416	Ehrlichia chaffeen
58	53.6	280	23	AAU06945	E. chafeensis OMP-
59	53.6	280	23	ABG77937	Ehrlichia chaffeen
60	53.6	280	23	AAU96107	Ehrlichia chafeens
61	53.6	280	23	AAU73414	Ehrlichia chaffeen
62	53.6	284	23	AAU96111	Cowdria ruminantiu
63	49	107	23	AAU17077	Human acyl CoA syn
64	49	132	22	AAU04201	Variable surface a
65	49	133	19	AAWS1097	Ehrlichia canis VS
66	49	133	21	AAB36191	Ehrlichia canis pa
67	49	133	21	AAU71480	Ehrlichia canis im
68	49	133	23	AAU96103	Ehrlichia canis p2
69	49	283	21	AAU71478	Ehrlichia canis im
70	49	283	23	AAU96101	Ehrlichia canis p2
71	47	65	22	AAU63107	Propionibacterium
72	45	40.9	240	21 AAG20226	Arabidopsis thalia
73	45	40.9	364	21 AAG47084	Arabidopsis thalia
74	45	40.9	365	21 AAG47111	Arabidopsis thalia
75	45	40.9	374	21 AAG47083	Arabidopsis thalia
76	45	40.9	375	21 AAG47110	Arabidopsis thalia
77	45	40.9	442	21 AAG47082	Arabidopsis thalia
78	45	40.9	443	21 AAG47109	Arabidopsis thalia
79	45	40.9	806	22 AAG92136	C glutamicum prote
80	44	40.0	139	22 AAM24384	Human EST encoded
81	44	40.0	308	22 ABB58819	Drosophila melanog
82	44	40.0	1009	22 ABB66546	Drosophila melanog

83 43 39.1 131 22 AAM96181 Human reproductive
84 43 39.1 131 22 AAU18909 Novel prostate gla
85 43 39.1 215 22 AAU59863 Propionibacterium
86 43 39.1 280 23 ABU05952 M. tuberculosis an
87 43 39.1 904 20 AAY21976 Senescence-associat
88 43 39.1 1037 16 AAR75396 Flea sodium pump a
89 43 39.1 1648 23 ABB54925 Lactococcus lactis
90 42 38.2 111 21 AAY97821 Pseudomonas sp. WF
91 42 38.2 189 22 ABG10775 Novel human diagno
92 42 38.2 525 22 ABG24436 Novel human diagno
93 42 38.2 1763 22 ABB65838 Drosophila melanoq
94 41.5 37.7 368 24 ABP79011 N. gonorrhoeae ami
95 41.5 37.7 534 23 ABP64733 Human protein SEQ
96 41.5 37.7 622 23 ABA49033 Listeria monocytog
97 41.5 37.7 782 22 AAB93202 Human protein sequ
98 41 37.3 155 22 AAU00890 Human cancer relat
99 41 37.3 198 21 AAG07319 Arabidopsis thalia
100 41 37.3 198 21 AAG36298 Arabidopsis thalia

ALIGNMENTS

RESULT 1
ABG30744
ID ABG30744 standard; Peptide; 20 AA.
XX AC ABG30744;
DT 21-OCT-2002 (first entry)
XX DE Ehrlichia canis peptide fragment #2.
XX KW Antibody detection; monoclonal antibody; polyclonal antibody.
XX OS Ehrlichia canis.
XX PN WO200257794-A2.
XX PD 25-JUL-2002.
XX PF 16-JAN-2002; 2002WO-US01395.
XX PR 18-JAN-2001; 2001US-0765739.
XX PA (INDEX-) IDEXX LAB INC.
XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX DR WPI; 2002-599730/64.
XX PT New composition of matter comprising a polypeptide, useful in detecting
XX PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
XX PT detecting or quantifying the presence of Ehrlichia infection in mammals

Claim 1; Page 5; 29pp; English.

The invention relates to a composition of matter comprising a polypeptide isolated from Ehrlichia species. The composition can be used for detecting the presence of antibodies to Ehrlichia, comprising contacting one or more polypeptides with a test sample suspected of comprising antibodies to Ehrlichia, under conditions that allow polypeptide/antibody complexes to form and detecting the complexes, where the detection of polypeptide/antibody complexes is an indication that antibodies to Ehrlichia are present in the test sample. The composition is useful for detecting or quantifying the presence of E. canis or E. chaffeensis infection in mammals. The polypeptides can be used to develop monoclonal and/or polyclonal antibodies that can be employed in assay systems and the generation of chimeric antibodies for therapeutic use or other similar applications. This sequence represents an E. canis peptide fragment used in the composition of the invention.

SQ Sequence 20 AA;
Query Match 100.0%; Score 110; DB 23; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 NNTTGVFGLKQDWGATIKD 20
|||||
DB 1 NNTTGVFGLKQDWGATIKD 20
|||||
RESULT 2
AAY06959
ID AAY06959 standard; Protein; 288 AA.
XX AC AAY06959;
DT 05-JUL-1999 (first entry)
XX DE E. canis P30 protein.
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX OS detection; dog.
XX OS Ehrlichia canis.
XX PN WO9913720-A1.
XX PD 25-MAR-1999.
XX PF 18-SEP-1998; 98WO-US19600.
XX PR 19-SEP-1997; 97US-0059353.
XX PA (OHIS) UNIV OHIO STATE.
XX PI Ohashi N, Rikihisa Y;
XX DR WPI; 1999-254290/21.
XX DR N-PSDB; AAX34759.
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX PT Ehrlichia canis
XX PS Disclosure; Fig 19B; 55pp; English.
XX CC The invention provides isolated outer membrane proteins (OMP) from
XX CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX CC of the OMP family and consist of proteins OMP-1, -11B to -21, showing
XX CC in AAY06943-958. The E. canis proteins form part of the P30 family and
XX CC consist of proteins shown in AAY06959-970. The proteins and genes are
XX CC used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 288 AA;

Query Match 100.0%; Score 110; DB 20; Length 288;
Best Local Similarity 100.0%; Pred. No. 2.1e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGATIKD 20
|||||
DB 60 NNTTGVFGLKQDWGATIKD 79
|||||

RESULT 3

ABG77950
ID ABG77950 standard; Protein; 288 AA.
XX AC ABG77950;
XX DT 15-NOV-2002 (first entry)
XX DE Ehrlichia canis outer membrane protein (P30F) #1.

XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

XX OS Ehrlichia canis.

XX PN US2002120115-A1.

XX PD 29-AUG-2002.

XX PF 28-JAN-2002; 2002US-0059964.

XX PR 19-MAY-1999; 99US-0314701.

XX PA (RIKI/) RIKIHISA Y.

XX PA (OHAS/) OHASHI N.

XX PI Rikihisa Y, Ohashi N;

XX DR WPI; 2002-618954/66.

XX DR N-PSDB; ABS63291.

XX PT Isolated polynucleotide encoding an outer membrane protein of E.canis

XX PT or E.chaffeensis used in the diagnosis of infection -

XX PS Claim 10; Fig 19B; 49pp; English.

XX CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.

XX SQ Sequence 288 AA;

Query Match 100.0%; Score 110; DB 23; Length 288;

Best Local Similarity 100.0%; Pred. No. 2.1e-09;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGATIKD 20

|||||

Db 60 NTTGVFGLKQDWGATIKD 79

RESULT 4

ABG30749

ID ABG30749 standard; Peptide; 19 AA.

XX AC ABG30749;

XX DT 21-OCT-2002 (first entry)

XX DE Ehrlichia chaffeensis peptide fragment #5.

XX KW Antibody detection; monoclonal antibody; polyclonal antibody.

XX OS Ehrlichia chaffeensis.

XX PN WO200257794-A2.

XX PD 25-JUL-2002.

XX PF 16-JAN-2002; 2002WO-US01395.

XX PR 18-JAN-2001; 2001US-0765739.

XX PA (IDEX-) IDEX LAB INC.

XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;

XX

DR WPI; 2002-599730/64.

XX

PT New composition of matter comprising a polypeptide, useful in detecting
 the presence of antibodies to Ehrlichia canis or chaffeensis, or in
 detecting or quantifying the presence of Ehrlichia infection in mammals

XX

PS Claim 1; Page 5; 29pp; English.

XX

CC The invention relates to a composition of matter comprising a polypeptide
 isolated from Ehrlichia species. The composition can be used for
 detecting the presence of antibodies to Ehrlichia, comprising contacting
 one or more polypeptides with a test sample suspected of comprising
 antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 complexes to form and detecting the complexes, where the detection of
 polypeptide/antibody complexes is an indication that antibodies to
 Ehrlichia are present in the test sample. The composition is useful for
 detecting or quantifying the presence of E. canis or E. chaffeensis
 infection in mammals. The polypeptides can be used to develop monoclonal
 and/or polyclonal antibodies that can be employed in assay systems and in
 the generation of chimeric antibodies for therapeutic use or other
 similar applications. This sequence represents an E. chaffeensis peptide
 fragment used in the composition of the invention.

XX SQ Sequence 19 AA;

Query Match 87.3%; Score 96; DB 23; Length 19;

Best Local Similarity 94.4%; Pred. No. 1.8e-08;

Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWGATI 18

|||||

Db 1 NTTGVFGLKQDWGATI 18

RESULT 5

AAAY06948

ID AAAY06948 standard; Protein; 280 AA.

XX AC AAAY06948;

XX DT 05-JUL-1999 (first entry)

XX DE E. chaffeensis OMP-1F protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 XX detection; dog.

XX OS Ehrlichia chaffeensis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX DR WPI; 1999-254290/21.

XX DR N-PSDB; AAX34748.

XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 XX Ehrlichia canis

XX PS Claim 16; Fig 8B; 55pp; English.

XX

CC The invention provides isolated outer membrane proteins (OMP) from
 Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 280 AA;

Query Match 87.3%; Score 96; DB 20; Length 280;
Best Local Similarity 94.4%; Pred. No. 3.6e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGATI 18
|||||
Db 60 NTTTGVFLKQDWGATI 77

RESULT 6
ABG777940
ID ABG77940 standard; Protein; 280 AA.
XX
AC ABG77940;
DT
DT 15-NOV-2002 (first entry)
XX
XX Ehrlichia chaffeensis outer membrane protein (OMP) #6.
DE Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
KW
XX Ehrlichia chaffeensis.
OS
PN US2002120115-A1.
XX
XX 29-AUG-2002.
XX
XX 28-JAN-2002; 2002US-0059964.
XX
XX 19-MAY-1999; 99US-0314701.
XX
XX (RIKI/) RIKIHISA Y.
FA (OHAS/) OHASHI N.
PI Rikihisa Y, Ohashi N;
PI
DR WPI; 2002-618954/66.
DR N-PSDB; ABS63281.
XX
XX Isolated polynucleotide encoding an outer membrane protein of E. canis
XX or E. Chaffeensis used in the diagnosis of infection -
XX Disclosure; Fig 8B; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX in the diagnosis of infection. An infection such as human ehrlichiosis or
XX canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX patient, providing a polypeptide or mixture of polypeptides, contacting
XX the sample with the polypeptide and assaying for the formation of a
XX complex between antibodies in the serum sample and the polypeptide where
XX formation of a complex is indicative of infection with E. chaffeensis.
XX This sequence represents an Ehrlichia outer membrane protein of the
XX invention.
XX
SQ Sequence 280 AA;

Query Match 87.3%; Score 96; DB 23; Length 280;
Best Local Similarity 94.4%; Pred. No. 3.6e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGATI 18
|||||
Db 60 NTTTGVFLKQDWGATI 77

RESULT 7
AAU96110
ID AAU96110 standard; Protein; 280 AA.
XX
XX AAU96110;
AC
DT 02-JUL-2002 (first entry)
XX
XX Ehrlichia chaffeensis OMP-1F.
DE
XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
KW antibacterial.
XX
XX Ehrlichia chaffeensis.
OS
XX WO200222782-A2.
PN
XX
XX 21-MAR-2002.
PD
XX
XX 12-SEP-2001; 2001WO-US28759.
PF
XX
XX 12-SEP-2000; 2000US-0660587.
PR
XX
XX (RERE-) RES DEV FOUND.
PA
XX Walker DH, Yu X, McBride JW;
XX
XX WPI; 2002-351882/38.
DR
XX
XX New recombinant homologous 28 kilodalton immunodominant protein from
XX Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX
XX Example 3; Figure 3; 106pp; English.
XX
XX The invention relates to a recombinant homologous 28 kDa immunodominant
XX protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
XX dispersed in a pharmaceutically acceptable carrier, is useful for
XX inhibiting E. canis infection in a subject. (I) is useful in the
XX development of vaccines and serodiagnostics that are particularly
XX effective for disease prevention and serodiagnosis. AAU96100-AAU96118
XX represent the 28-kDa antigen amino acid sequences of the invention.
XX
SQ Sequence 280 AA;

Query Match 87.3%; Score 96; DB 23; Length 280;
Best Local Similarity 94.4%; Pred. No. 3.6e-07;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGATI 18
|||||
Db 60 NTTTGVFLKQDWGATI 77

RESULT 8
AAW51094
ID AAW51094 standard; Protein; 280 AA.
XX
XX AAW51094;
AC
XX
XX 14-SEP-1998 (first entry)
DT
XX
XX Ehrlichia chaffeensis VSA4 protein.
DE
XX
XX MAP1 homologue; variable surface antigen; VSA4; rickettsia;
KW DNA vaccine.
XX
XX Ehrlichia chaffeensis.
OS
XX
XX Key Location/Qualifiers
FT Peptide 1..25
FT /note= "putative signal peptide"
XX
XX WO9816554-A1.
PN

XX PD 23-APR-1998.
 XX PE 17-OCT-1997; 97WO-US19044.
 XX PR 17-OCT-1996; 96US-0733230.
 XX (UYFL) UNIV FLORIDA.
 XX PI Barbet AF, Burr ridge MJ, Ganta RR, Mahan SM, McGuire TC;
 XX PI Nyika A, Rurangirwa FR;
 XX DR WPI; 1998-251232/22.
 XX DR N-PSDB; AAV07179.
 XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX Claim 3; Fig 2B; 39pp; English.
 XX This is the full-length variable surface antigen VSA4 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX SQ Sequence 280 AA;

Query Match 81.8%; Score 90; DB 19; Length 280;
 Best Local Similarity 88.9%; Pred. No. 3.3e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQWDGATI 18
 ||| ||||| ||||| |||
 Db 60 NTTTGVFLKQWDGSTI 77

RESULT 9
 AAB36188
 ID AAB36188 standard; Protein; 280 AA.
 XX
 XX AC AAB36188;
 XX DT 02-MAR-2001 (first entry)
 XX DE Ehrlichia chaffeensis partial VSA4.
 XX KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX OS Ehrlichia chaffeensis.
 XX PN WC200005063-A2.
 XX PD 02-NOV-2000.
 XX PF 21-APR-2000; 2000WO-US10886.
 XX PR 22-APR-1999; 99US-0130725.
 XX (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Bowie MV, Ganta RR, Burr ridge MJ, Mahan SM, McGuire TC;
 XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 XX DR WPI; 2000-679675/66.
 XX DR N-PSDB; AAC68705.
 XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 Claim 3; Page 45-46; 63pp; English.
 XX The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccine to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX SQ Sequence 280 AA;

Query Match 81.8%; Score 90; DB 21; Length 280;
 Best Local Similarity 88.9%; Pred. No. 3.3e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQWDGATI 18
 ||| ||||| ||||| |||
 Db 60 NTTTGVFLKQWDGSTI 77

RESULT 10
 AAU04198
 ID AAU04198 standard; Protein; 280 AA.
 XX
 XX AC AAU04198;
 XX DT 23-OCT-2001 (first entry)
 XX DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
 XX KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX OS Ehrlichia chaffeensis.
 XX PN US6251872-B1.
 XX PD 26-JUN-2001.
 XX PF 17-OCT-1997; 97US-0953326.
 XX PR 17-OCT-1996; 96US-0733230.
 XX (UYFL) UNIV FLORIDA.
 XX PI Barbet AF, Ganta RR, McGuire TC, Burr ridge MJ, Nyika A;
 XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX DR WPI; 2001-424487/45.
 XX DR N-PSDB; AAS07578.
 XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures
 XX

PS Example 3; Fig 2A-2B; 30pp; English.

XX The sequence represents the amino acid sequence of variable surface
CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.

XX Sequence 280 AA;

Query Match 81.8%; Score 90; DB 22; Length 280;

Best Local Similarity 88.9%; Pred. No. 3.3e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDGATI 18
||| ||||| ||||| |||
Db 60 NTTGVFGLKQDWDGATI 77

RESULT 11

ID AAU73417 standard; Protein; 280 AA.

XX AAU73417;

DT 12-MAR-2002 (first entry)

DE Ehrlichia chaffeensis outer membrane protein P28-18.

KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.

OS Ehrlichia chaffeensis.

PN WO200183699-A2.

PD 08-NOV-2001.

PF 01-MAY-2001; 2001WO-US13997.

PR 01-MAY-2000; 2000US-201035P.

PA (RERE-) RES DEV FOUND.

PI Walker DH, Yu X;

DR WPI; 2002-086527/09.

PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis -

PS Disclosure; Figure 2; 97pp; English.

XX The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.

XX Sequence 280 AA;

Query Match 81.8%; Score 90; DB 23; Length 280;

Best Local Similarity 88.9%; Pred. No. 3.3e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDGATI 18
||| ||||| ||||| |||
Db 60 NTTGVFGLKQDWDGATI 77

RESULT 12

ID ABG30745 standard; Peptide; 20 AA.

XX ABG30745;

AC ABG30745;

DT 21-OCT-2002 (first entry)

DE Ehrlichia chaffeensis peptide fragment #1.

KW Antibody detection; monoclonal antibody; polyclonal antibody.

OS Ehrlichia chaffeensis.

PN WO200257794-A2.

PD 25-JUL-2002.

PF 16-JAN-2002; 2002WO-US01395.

PR 18-JAN-2001; 2001US-0765739.

PA (IDEX-) IDEXX LAB INC.

PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
DR WPI; 2002-599730/64.

PT New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals

XX Claim 1; Page 5; 29pp; English.

XX The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. chaffeensis peptide
CC fragment used in the composition of the invention.

SQ Sequence 20 AA;

Query Match 74.5%; Score 82; DB 23; Length 20;

Best Local Similarity 70.0%; Pred. No. 3.3e-06;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDGATIKD 20
||| ||||| ||||| |||

Db 1 NTTGVFGLKQDWDGSAISN 20

RESULT 13

AAU06942 standard; Protein; 256 AA.

XX AAU06942;

AC AAU06942;

DT 05-JUL-1999 (first entry)
 XX E. chafeensis p28 protein.
 DE
 XX
 KW Outer membrane protein; OMP; Ehrlichia chafeensis; E. canis; p30;
 KW detection; dog.
 XX
 XX Ehrlichia chafeensis.
 OS
 XX
 PN WO9913720-A1.
 XX
 XX 25-MAR-1999.
 PD
 XX 18-SEP-1998; 98WO-US19600.
 XX
 XX 19-SEP-1997; 97US-0059353.
 PR
 XX (OHIS) UNIV OHIO STATE.
 PA
 XX Ohashi N, Rikihisa Y;
 PI
 XX WPI; 1999-254290/21.
 DR N-PSDB; AAX34742.
 XX
 XX Novel outer membrane proteins from Ehrlichia chafeensis and
 PT Ehrlichia canis
 XX
 PS Claim 18; Fig 1; 55pp; English.
 XX
 CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chafeensis and E. canis. The E. chafeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chafeensis in patients and E. canis in dogs.
 CC
 XX Sequence 256 AA;
 SQ

Query Match 74.5%; Score 82; DB 20; Length 256;
 Best Local Similarity 70.0%; Pred. No. 5.7e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGATIKD 20
 III IIIIIIIIIII: I :
 Db 34 NTTGVFGLKQNDGSAISN 53

RESULT 14
 ABG77966
 ID ABG77966 standard; Protein; 256 AA.
 AC ABG77966;
 XX
 XX 15-NOV-2002 (first entry)
 DT
 DE Protein encoded by Ehrlichia chafeensis p28 gene.
 XX
 XX Outer membrane protein; OMP; p30F; ehrlichiosis; infection; p28; OMP-1.
 KW Ehrlichia chafeensis.
 XX
 OS US2002120115-A1.
 XX
 PN 29-AUG-2002.
 PD
 XX 28-JAN-2002; 2002US-0059964.
 PF
 XX 19-MAY-1999; 99US-0314701.
 PR
 XX (RIKL/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.
 XX
 XX Rikihisa Y, Ohashi N;
 PI

XX WPI; 2002-618954/66.
 DR N-PSDB; ABS63307.
 XX
 XX Isolated polynucleotide encoding an outer membrane protein of E. canis
 PT or E. chafeensis used in the diagnosis of infection -
 PT
 XX Disclosure; Fig 1; 49pp; English.
 PS
 XX
 CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chafeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chafeensis.
 CC This sequence represents the Ehrlichia chafeensis OMP-1 protein encoded
 CC by the p28 gene.
 XX
 XX Sequence 256 AA;
 SQ

Query Match 74.5%; Score 82; DB 23; Length 256;
 Best Local Similarity 70.0%; Pred. No. 5.7e-05;
 Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGATIKD 20
 III IIIIIIIIIII: I :
 Db 34 NTTGVFGLKQNDGSAISN 53

RESULT 15
 AAW51095
 ID AAW51095 standard; Protein; 276 AA.
 XX
 XX AAW51095;
 AC
 XX
 XX 14-SEP-1998 (first entry)
 DT
 XX Ehrlichia chafeensis VSA5 protein (partial sequence).
 DE
 XX MAP1 homologue; variable surface antigen; VSA5; rickettsia;
 KW DNA vaccine.
 KW
 XX Ehrlichia chafeensis.
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /note= "putative signal peptide"
 XX
 PN WO9816554-A1.
 XX
 XX 23-APR-1998.
 PD
 XX 17-OCT-1997; 97WO-US19044.
 XX
 XX 17-OCT-1996; 96US-0733230.
 PR
 XX (UYFL) UNIV FLORIDA.
 PA
 XX Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 PI
 XX WPI; 1998-251232/22.
 DR N-PSDB; AAV07179.
 DR
 XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 XX Claim 3; Fig 2B; 39pp; English.
 PS
 XX This is the near full-length variable surface antigen VSA5 protein
 CC

Matches	14;	Conservative	3;	Mismatches	3;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

```

QY      1 NTTTGVFGLKQDWDGATIKD 20
        { | | | | | | | | | | :
Db     59 NTTTGVFGLKQDWDGSAISN 78

```

RESULT 18
AAW51089
ID AAW51089 standard; Protein: 280 AA.

AA
AC AAW51089;XX
DT 14-SEP-1998 (first entry)

XX
DE
Ehrlichia chaffeensis major antigenic protein 1 (MAP1)

KW MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.

XX OS Ehrlichia chaffeensis.

[illegible]XX
PD
23-APR-1998

XX
PF 17-OCT-1997: 97WQ-IIS1904A

XX
PR 17-OCT-1996: 96US-0733230XX
PA (UYFL) UNTV FLORIDA

XX
PI Barbet AF, Burrledge MJ, Ganta RR, Mahan SM, McGuire TC.

PI Nýika A, Rurangirwa FR;
XX

DR WPI; 1998-251232/
DR N-PSDB: AAV07177.

Composition containing nucleic acid encoding rickettsial antigen -
 useful for, e.g. stimulating protective immune response in humans or
 animals

XX
PS
Claim 3: Page 18-19: 3900: English

This polypeptide comprises the major antigen protein 1 gene (MAP1) of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see AAV071177). A claimed composition comprises a nucleic acid (see AAV071176-82) encoding a polypeptide (see AAW51088-99) that elicits a protective immune response against a rickettsial pathogen. The nucleic acid is used, in human or veterinary medicine, in vaccines to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The nucleic acid does not replicate in the host but remains episomal and capable of expressing polypeptide for at least 19 mth. The Ehrlichia antigenic polypeptides can be used diagnostically to detect antibodies associated with Ehrlichia infection (claimed).

XX	Sequence	280 AA:
SQ		

Query Match	74.5%;	Score 82;	DB 19;	Length 280;
Best Local Similarity	70.0%;	Pred. NO. 6.3e-05;		
Matches	14.	Conservative	3.	Mismatches
				3.
				Indels

QY 1 NTTTGVFGLKQDWDGATIKD 20
||| ||||| :
DB 50 NTTTGVFGLKQDWDGATIKD 70

RESULT 19
AAB36183
ID: AAB36183 standard: Prot. 01. 280. 11

XX
AC
AAB26103.

XX

PN US6251872-B1.
XX 26-JUN-2001.
XX
XX PF 17-OCT-1997; 97US-0953326.
XX PR 17-OCT-1996; 96US-0733230.
XX (UYFL) UNIV FLORIDA.
XX Barbet AF, Ganta RR, McGuire TC, Burr ridge MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX WPI: 2001-424487/45.
XX N-PSDB; AAS07576.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures -
XX
XX Disclosure: Column 15-17; 30pp; English.
XX
XX The sequence represents the amino acid sequence of major antigenic
CC protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
CC polypeptides are useful as vaccines for conferring immunity to rickettsia
CC infection, including Cowdria ruminantium causing heartwater. The MAP
CC polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.
XX
XX Sequence 280 AA;
XX
XX Query Match 74.5%; Score 82; DB 22; Length 280;
XX Best Local Similarity 70.0%; Pred. NO. 6.3e-05;
XX Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NTTTGVFGLKQWDGATIKD 20
DB III IIIIIIIIIII :
60 NTTTGVFGLKQWDGSAISN 79
RESULT 21
ID AAY06943 standard; Protein; 281 AA.
XX
XX AAY06943;
XX
XX 05-JUL-1999 (first entry)
XX
XX E. chaffeensis OMP-1 protein.
XX
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.
XX
XX Ehrlichia chaffeensis.
XX
XX WO9913720-A1.
XX
XX 25-MAR-1999.
XX
XX PF 18-SEP-1998; 98WO-US19600.
XX
XX PR 19-SEP-1997; 97US-0059353.
XX
XX (OHIS) UNIV OHIO STATE.
XX
XX Ohashi N, Rikihisa Y;
PI

XX WPI: 1999-254290/21.
DR N-PSDB; AAX34743.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
XX Disclosure: Fig 3B; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06999-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 281 AA;
XX
XX Query Match 74.5%; Score 82; DB 20; Length 281;
XX Best Local Similarity 70.0%; Pred. NO. 6.3e-05;
XX Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NTTTGVFGLKQWDGATIKD 20
DB III IIIIIIIIIII :
59 NTTTGVFGLKQWDGSAISN 78
RESULT 22
ID ABG77935 standard; Protein; 281 AA.
XX
XX ABG77935;
XX
XX 15-NOV-2002 (first entry)
XX
XX Ehrlichia chaffeensis outer membrane protein (OMP) #1.
DE
XX
XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
XX Ehrlichia chaffeensis.
XX
XX US2002120115-A1.
XX
XX 29-AUG-2002.
XX
XX 28-JAN-2002; 2002US-0059964.
XX
XX 19-MAY-1999; 99US-0314701.
XX
XX (RIKI/) RIKIHISA Y.
XX (OHAS/) OHASHI N.
XX
XX Rikihisa Y, Ohashi N;
XX
XX WPI: 2002-618954/66.
XX N-PSDB; ABS63276.
XX
XX Isolated polynucleotide encoding an outer membrane protein of E. canis
PT or E. chaffeensis used in the diagnosis of infection -
XX
XX Claim 14; Fig 3B; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX

SQ Sequence 281 AA;

Query Match 74.5%; Score 82; DB 23; Length 281;
Best Local Similarity 70.0%; Pred. No. 6.3e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVFGLKQWDGATIKD 20
Db 59 NTTGVFGLKQWNGSAISN 78
||| |||||:||||: | :

RESULT 23
AAU96105
ID AAU96105 standard; Protein; 281 AA.
XX AC AAU96105;
XX DT 02-JUL-2002 (first entry)
XX DE Ehrlichia chaffeensis P28.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX OS Ehrlichia chaffeensis.
XX PN WO200222782-A2.
XX PD 21-MAR-2002.
XX PF 12-SEP-2001; 2001WO-US28759.
XX PR 12-SEP-2000; 2000US-0660587.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X, McBride JW;
XX DR WPI; 2002-351882/38.
XX PT New recombinant homologous 28 kilodalton immunodominant protein from Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX PS Example 3; Figure 3; 106pp; English.
XX CC The invention relates to a recombinant homologous 28 kDa immunodominant protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably dispersed in a pharmaceutically acceptable carrier, is useful for inhibiting E. canis infection in a subject. (I) is useful in the development of vaccines and serodiagnostics that are particularly effective for disease prevention and serodiagnosis. AAU96100-AAU96118 represent the 28-kDa antigen amino acid sequences of the invention.
XX SQ Sequence 281 AA;

Query Match 74.5%; Score 82; DB 23; Length 281;
Best Local Similarity 70.0%; Pred. No. 6.3e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVFGLKQWDGATIKD 20
Db 59 NTTGVFGLKQWNGSAISN 78
||| |||||:||||: | :

RESULT 24
AAU73418
ID AAU73418 standard; Protein; 281 AA.
XX AC AAU73418;
XX DT 12-MAR-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein P28-19.

XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX Ehrlichia chaffeensis.
XX WO200183699-A2.
XX 08-NOV-2001.
XX 01-MAY-2001; 2001WO-US13997.
XX 01-MAY-2000; 2000US-201035P.
XX (RERE-) RES DEV FOUND.
XX Walker DH, Yu X;
XX WPI; 2002-066527/09.
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated P28 useful as a vaccine against Ehrlichia chaffeensis -
XX Disclosure; Figure 2; 97pp; English.
XX The invention relates to isolated and purified 28-kDa outer membrane proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins are encoded by a 28kDa outer membrane protein multigene family. P28 proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28 is useful for transfecting a host cell. AAU73400-AAU73420 represent Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX SQ Sequence 281 AA;

Query Match 74.5%; Score 82; DB 23; Length 281;
Best Local Similarity 70.0%; Pred. No. 6.3e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVFGLKQWDGATIKD 20
Db 59 NTTGVFGLKQWNGSAISN 78
||| |||||:||||: | :

RESULT 25
AAU06962
ID AAU06962 standard; Protein; 280 AA.
XX AC AAU06962;
XX 05-JUL-1999 (first entry)
XX E. canis P30-2 protein.
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30; detection; dog.
XX Ehrlichia canis.
XX WO9913720-A1.
XX 25-MAR-1999.
XX 18-SEP-1998; 98WO-US19600.
XX 19-SEP-1997; 97US-0059353.
XX (OHIS) UNIV OHIO STATE.
XX Ohashi N, Rikihisa Y;
XX WPI; 1999-254290/21.
XX N-PSDB; AAX34762.
XX Novel outer membrane proteins from Ehrlichia chaffeensis and

```

PT Ehrlichia canis
XX Disclosure; Fig 22B; 55pp; English.
PS
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 280 AA;
Query Match 70.0%; Score 77; DB 20; Length 280;
Best Local Similarity 65.0%; Pred. No. 0.0004;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGATIKD 20
DB 59 NSTVGVFGLKHDWNGGTISN 78

RESULT 26
ID AAY71479 standard; Protein; 280 AA.
XX
AC AAY71479;
XX
DT 12-OCT-2000 (first entry)
XX
DE Ehrlichia canis immunoreactive protein Eca28SA3.
XX
KW Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;
KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
KW tick-borne rickettsial disease; serodiagnosis.
XX
OS Ehrlichia canis.
XX
EH Key Location/Qualifiers
FT Peptide 1..23
FT Protein /label= Signal_peptide
FT Protein 24..280
FT Protein /label= Mature_Eca28SA3_28-kDa_protein
XX
PN WO200032745-A2.
XX
PD 08-JUN-2000.
XX
PF 24-NOV-1999; 99WO-US28075.
XX
PR 30-NOV-1998; 98US-0201458.
XX
PR 03-MAR-1999; 99US-0261358.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X, McBride JW;
XX
PI WPI: 2000-412298/35.
XX
DR N-PSDB; AAD01294, AAD01295.
XX
PT Ehrlichia canis antigens useful for vaccinating against canine
PT ehrlichiosis in dogs -
XX
PS Claim 12; Page 68-69; 86pp; English.
XX
CC The patent relates to homologous 28-kilodalton (kDa) protein genes of
CC Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and
CC Eca28-2. These genes are members of a polymorphic multiple gene family
CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are
CC immunoreactive with anti-E. canis serum hence are important
CC immunoprotective antigens. The protein is useful for vaccinating
CC against E. canis infections such as canine ehrlichiosis in dogs.
XX

```

```

CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
CC tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst
CC different strains of E. canis and hence useful for serodiagnosis of
CC canine ehrlichiosis. The present sequence is a E. canis
CC Eca28SA3 30-kDa protein which is post-translationally modified to a
CC mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX
SQ Sequence 280 AA;
Query Match 70.0%; Score 77; DB 21; Length 280;
Best Local Similarity 65.0%; Pred. No. 0.0004;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGATIKD 20
DB 59 NSTVGVFGLKHDWNGGTISN 78

RESULT 27
ID ABG77953 standard; Protein; 280 AA.
XX
AC ABG77953;
XX
DT 15-NOV-2002 (first entry)
XX
DE Ehrlichia canis outer membrane protein (P30F) #4.
XX
KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
OS Ehrlichia canis.
XX
PN US2002120115-A1.
XX
PD 29-AUG-2002.
XX
PF 28-JAN-2002; 2002US-0059964.
XX
PR 19-MAY-1999; 99US-0314701.
XX
PA (RIKI/) RIKIHISA Y.
PA (OHAS/) OHASHI N.
XX
PI Rikihisa Y, Ohashi N;
XX
WPI: 2002-618954/66.
XX
DR N-PSDB; ABS63294.
XX
PT Isolated polynucleotide encoding an outer membrane protein of E. canis
PT or E. chaffeensis used in the diagnosis of infection -
XX
PS Claim 10; Fig 22B; 49pp; English.
XX
CC The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX
SQ Sequence 280 AA;
Query Match 70.0%; Score 77; DB 23; Length 280;
Best Local Similarity 65.0%; Pred. No. 0.0004;
Matches 13; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGATIKD 20
DB 59 NSTVGVFGLKHDWNGGTISN 78

```

```

PD      25-MAR-1999.
XX
XX      18-SEP-1998;   98WO-US19600.
XX
XX      19-SEP-1997;   97US-0059353.
XX
XX      (OHIS ) UNIV OHIO STATE.
XX
XX      Ohashi N, Rikihisa Y;
XX
XX      WPI: 1999-254290/21.
XX
XX      N-PSDB: AAX34764.
XX
XX
XX      Novel outer membrane proteins from Ehrlichia chaffeensis and
XX      Ehrlichia canis
XX
XX      Disclosure; Fig 24B; 55pp; English.
XX
XX
XX      The invention provides isolated outer membrane proteins (OMP) from
XX      Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX      of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX      in AAU06943-958. The E. canis proteins form part of the p30 family and
XX      consist of proteins shown in AAU06959-970. The proteins and genes are
XX      used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX      Sequence 276 AA;
XX
XX      Query Match          69.1%; Score 76; DB 20; Length 276;
XX      Best Local Similarity 65.0%; Pred. No. 0.00056;
XX      Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
XX
QY      1 NTGTGVLKQDWDGATIKD 20
Db      |||||:||||: | | |
        60 NTTTGIFGLKESWTGGIILD 79

RESULT 30
ABG77955
ID      ABG77955 standard; Protein; 276 AA.
XX
XX      AC      ABG77955;
XX
XX      DT      15-NOV-2002 (first entry)
XX
XX      DE      Ehrlichia canis outer membrane protein (P30F) #6.
XX
XX      KW      Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
XX      OS      Ehrlichia canis.
XX
XX      PN      US2002120115-A1.
XX
XX      PD      29-AUG-2002.
XX
XX      PF      28-JAN-2002; 2002US-0059964.
XX
XX      PR      19-MAY-1999; 99US-0314701.
XX
XX      PA      (RIKI/) RIKIHISA Y.
XX      (OHAS/) OHASHI N.
XX
XX      PI      Rikihisa Y, Ohashi N;
XX
XX      DR      WPI: 2002-618954/66.
XX      DR      N-PSDB: ABS63296.
XX
XX      Isolated polynucleotide encoding an outer membrane protein of E.canis
XX      or E.chaffeensis used in the diagnosis of infection -
XX
XX      Claim 10; Fig 24B; 49pp; English.
XX
XX      The invention relates to an isolated polynucleotide encoding an outer
XX      membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used

```

CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.
 XX
 XX SQ Sequence 276 AA;

Query Match 69.1%; Score 76; DB 23; Length 276;
 Best Local Similarity 65.0%; Pred. No. 0.00056;
 Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 NNTTGVFGLKQDWGATIKD 20
 Db 60 NNTTGFGLKESWTGGIILD 79

RESULT 31
 AAU96117
 ID AAU96117 standard; Protein: 276 AA.
 AC AAU96117;
 XX
 XX 02-JUL-2002 (first entry)
 DE Ehrlichia canis p28-3.
 XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
 KW antibacterial.
 XX Ehrlichia canis.
 OS
 XX WO200222782-A2.
 PN
 XX 21-MAR-2002.
 PD
 XX 12-SEP-2001; 2001WO-US28759.
 PF
 XX 12-SEP-2000; 2000US-0660587.
 PR
 XX (RERE-) RES DEV FOUND.
 PA
 XX Walker DH, Yu X, McBride JW;
 PI WPI: 2002-351882/38.
 DR N-PSDB; ABX68877.
 XX
 XX New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections
 PS
 XX Claim 16; Figure 15; 106pp; English.

CC The invention relates to a recombinant homologous 28 kDa immunodominant
 CC protein, p28, (I) of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AAU96110-AAU96118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX
 XX SQ Sequence 276 AA;

Query Match 69.1%; Score 76; DB 23; Length 276;
 Best Local Similarity 65.0%; Pred. No. 0.00056;
 Matches 13; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 NNTTGVFGLKQDWGATIKD 20
 Db 60 NNTTGFGLKESWTGGIILD 79

RESULT 32
 ABG30747
 ID ABG30747 standard; Peptide: 19 AA.
 AC ABG30747;
 XX
 XX 21-OCT-2002 (first entry)
 DT
 XX Ehrlichia chaffeensis peptide fragment #3.
 DE
 XX Antibody detection; monoclonal antibody; polyclonal antibody.
 KW
 XX Ehrlichia chaffeensis.
 OS
 XX WO200257794-A2.
 PN
 XX 25-JUL-2002.
 PD
 XX 16-JAN-2002; 2002WO-US01395.
 PF
 XX 18-JAN-2001; 2001US-0765739.
 PR
 XX (IDEX-) IDEXX LAB INC.
 PA
 XX Lawton R, O'Connor TP, Bartol BA, Machenry PS;
 PI WPI: 2002-599730/64.
 DR
 XX
 XX New composition of matter comprising a polypeptide, useful in detecting
 PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
 PT detecting or quantifying the presence of Ehrlichia infection in mammals
 PT
 XX
 XX Claim 1; Page 5; 29pp; English.

CC The invention relates to a composition of matter comprising a polypeptide
 CC isolated from Ehrlichia species. The composition can be used for
 CC detecting the presence of antibodies to Ehrlichia, comprising contacting
 CC one or more polypeptides with a test sample suspected of comprising
 CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 CC complexes to form and detecting the complexes. Here the detection of
 CC polypeptide/antibody complexes is an indication that antibodies to
 CC Ehrlichia are present in the test sample. The composition is useful for
 CC detecting or quantifying the presence of E. canis or E. chaffeensis
 CC infection in mammals. The polypeptides can be used to develop monoclonal
 CC and/or polyclonal antibodies that can be employed in assay systems and in
 CC the generation of chimeric antibodies for therapeutic use or other
 CC similar applications. This sequence represents an E. chaffeensis peptide
 CC fragment used in the composition of the invention.
 XX
 XX SQ Sequence 19 AA;

Query Match 64.5%; Score 71; DB 23; Length 19;
 Best Local Similarity 66.7%; Pred. No. 0.00018;
 Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NNTTGVFGLKQDWGAT 18
 Db 1 NNTTGVFGLKQDWGAT 18

RESULT 33
 AAWS1092
 ID AAWS1092 standard; Protein: 286 AA.
 AC AAWS1092;
 XX
 XX 14-SEP-1998 (first entry)
 DT
 XX Ehrlichia chaffeensis VSA2 protein.
 DE
 XX MAP1 homologue; variable surface antigen; VSA2; rickettsia;

KW DNA vaccine.
 XX Ehrlichia chaffeensis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /note= "putative signal peptide"
 XX
 XX W09816554-AL.
 XX
 XX 23-APR-1998.
 XX
 XX 17-OCT-1997; 97WO-US19044.
 XX
 XX 17-OCT-1996; 96US-0733230.
 XX
 XX (UYFL) UNIV FLORIDA.
 XX
 XX Barbet AF, Burrigide MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 PI
 XX WPI; 1998-251232/22.
 XX N-PSDB; AAV07179.
 XX
 XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 PT
 XX
 XX Claim 3; Fig 2A; 39pp; English.
 XX
 XX This is the full-length variable surface antigen VSA2 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF2) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 XX Sequence 286 AA;
 SQ
 Query Match 64.5%; Score 71; DB 19; Length 286;
 Best Local Similarity 66.7%; Pred. No. 0.0037;
 Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NTTTGVFLKQDWGATI 18
 ||| ||||:|||| |
 Db 60 NTTVGFGIEQDWDRCVI 77
 RESULT 34
 AAY06946
 ID AAY06946 standard; Protein; 286 AA.
 XX
 XX AAY06946;
 AC
 XX
 XX 05-JUL-1999 (first entry)
 DT
 XX E. chaffeensis OMP-1D protein.
 DE
 XX
 XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX
 XX Ehrlichia chaffeensis.
 OS
 XX W09913720-AL.
 PN
 XX

PD 25-MAR-1999.
 XX
 XX 18-SEP-1998; 98WO-US19600.
 XX
 XX 19-SEP-1997; 97US-0059353.
 XX
 XX (OHIS) UNIV OHIO STATE.
 XX
 XX Ohashi N, Rikihisa Y;
 PI
 XX WPI; 1999-254290/21.
 DR N-PSDB; AAX34746.
 DR
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 PT
 XX Claim 14; Fig 6B; 55pp; English.
 PS
 XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX
 XX Sequence 286 AA;
 SQ
 Query Match 64.5%; Score 71; DB 20; Length 286;
 Best Local Similarity 66.7%; Pred. No. 0.0037;
 Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 NTTTGVFLKQDWGATI 18
 ||| ||||:|||| |
 Db 60 NTTVGFGIEQDWDRCVI 77
 RESULT 35
 AAB36186
 ID AAB36186 standard; Protein; 286 AA.
 XX
 XX AAB36186;
 AC
 XX
 XX 02-MAR-2001 (first entry)
 DT
 XX Ehrlichia chaffeensis partial VSA2.
 DE
 XX
 XX Ehrlichia chaffeensis; VSA2; variable surface antigen 2; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX
 XX Ehrlichia chaffeensis.
 OS
 XX W0200065063-A2.
 PN
 XX
 XX 02-NOV-2000.
 PD
 XX 21-APR-2000; 2000WO-US10886.
 PF
 XX 22-APR-1999; 99US-0130725.
 PR
 XX (UYFL) UNIV FLORIDA.
 PA
 XX Barbet AF, Bowie MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Allemen AR;
 PI
 XX WPI; 2000-679675/66.
 DR N-PSDB; AAC68703.
 DR
 XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 PT

Search completed: October 6, 2003, 07:47:56
Job time : 36.3731 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 10.209 Seconds
(without alignments)
178.981 Million cell updates/sec

Title: US-09-765-739A-5

Perfect score: 105

Sequence: 1 NTTVGVFGIEQDWRGVIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_76:*

1: piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	286	2 JE0219	28k surface antige
2	77	73.3	280	2 JE0217	28k surface antige
3	73	69.5	276	2 JE0218	28k surface antige
4	51	48.6	278	2 JE0216	28k surface antige
5	47	44.8	160	2 A75466	2-Demethylmenaquin
6	46	43.8	284	2 I40882	major antigenic pr
7	45	42.9	253	2 B95284	probable [imported
8	44.5	42.4	2182	1 GNNYB1	genome polypeptid
9	43	41.0	498	2 B90456	hypothetical prote
10	43	41.0	5188	2 B05547	probable RTX fami
11	43	41.0	5291	2 F90696	hypothetical prote
12	42.5	40.5	417	2 E84430	probable inositol
13	42	40.0	244	2 E84885	hypothetical prote
14	41	39.0	142	2 F83359	hypothetical prote
15	41	39.0	279	2 A11882	nitrate transport
16	41	39.0	299	2 S68198	probable transcrip
17	41	39.0	390	2 A81656	cystathionine beta
18	41	39.0	390	2 AG1284	cystathionine beta
19	41	39.0	415	2 D87020	probable membrane
20	41	39.0	573	2 T07948	mandelonitrile lya
21	41	39.0	576	2 T08073	mandelonitrile lya
22	41	39.0	688	2 JC2248	polypeptid - swee
23	40.5	38.6	205	2 A72479	hypothetical prote
24	40.5	38.6	259	2 E95357	probable [imported
25	40.5	38.6	319	2 C97199	activator of 2-hyd
26	40.5	38.6	428	2 A80471	probable DEAD-box
27	40.5	38.6	1790	1 S27772	vitellogenin precu
28	40	38.1	46	1 TZA21	toxin I - snake-lo
29	40	38.1	139	2 AD2997	hypothetical prote

30	40	38.1	141	2 E98286	hypothetical prote
31	40	38.1	143	1 H5YTA	ferredoxin [3Fe-4S
32	40	38.1	231	1 H95187	ribulose-phosphate
33	40	38.1	256	2 A10406	probable dehydroge
34	40	38.1	315	2 A90030	hypothetical prote
35	40	38.1	349	2 F70357	lipoprotein - Aquil
36	40	38.1	361	2 C85430	MAP kinase like pr
37	40	38.1	398	2 C83499	probable transglyc
38	40	38.1	436	2 T28066	hypothetical prote
39	40	38.1	530	2 T23255	hypothetical prote
40	40	38.1	540	2 S63299	sugar transport pr
41	40	38.1	781	2 G96991	secreted protease
42	40	38.1	1012	2 T42385	alpha-mannosidase
43	40	38.1	1090	2 C86577	PBP2-transglycolas
44	40	38.1	1090	2 D72048	PBP2-transglycolas
45	39.5	37.6	116	2 T36999	probable transposa
46	39	37.1	91	2 A11982	hypothetical prote
47	39	37.1	212	2 C90905	hypothetical prote
48	39	37.1	216	2 B85712	unknown protein en
49	39	37.1	265	2 T43123	probable positive
50	39	37.1	279	1 S56642	nitrate transport
51	39	37.1	322	2 H90600	prolipoprotein dia
52	39	37.1	331	2 A70884	probable ribF prot
53	39	37.1	331	2 T44902	probable riboflavi
54	39	37.1	359	2 AB0844	membrane-bound lyt
55	39	37.1	361	2 A65050	membrane-bound lyt
56	39	37.1	361	2 F91073	membrane-bound lyt
57	39	37.1	361	2 B85918	membrane-bound lyt
58	39	37.1	367	2 H83088	membrane-bound lyt
59	39	37.1	370	2 AG0359	probable membrane-
60	39	37.1	371	2 D69253	conserved hypotet
61	39	37.1	377	2 F69008	acetyltransferase
62	39	37.1	397	2 I58099	gene P2X3 protein
63	39	37.1	397	2 S60334	purinoceptor P2X -
64	39	37.1	413	2 AC1045	probable permease
65	39	37.1	462	2 D84751	hypothetical prote
66	39	37.1	503	2 T40997	probable short-cha
67	39	37.1	548	2 E72457	probable phenylala
68	39	37.1	574	2 T50766	mandelonitrile lya
69	39	37.1	751	2 C88485	protein F23F12.5 [
70	39	37.1	759	2 D81657	general secretion
71	39	37.1	871	2 T28706	hypothetical prote
72	39	37.1	899	2 G96617	probable disease r
73	39	37.1	900	2 H96617	probable disease r
74	39	37.1	951	2 T00260	hypothetical prote
75	39	37.1	993	2 J00488	afsr protein - Str
76	39	37.1	1098	2 S38100	hypothetical prote
77	39	37.1	1474	2 F69009	probable membrane
78	38.5	36.7	223	2 B70372	flagellar protein
79	38.5	36.7	267	2 C83242	conserved hypotet
80	38.5	36.7	317	2 AB3783	hypothetical prote
81	38.5	36.7	319	2 AE2294	hypothetical prote
82	38.5	36.7	453	2 T20965	hypothetical prote
83	38.5	36.7	487	2 T05127	hypothetical prote
84	38.5	36.7	644	2 S72827	probable arginyl-t
85	38	36.2	63	2 S72827	probable insertion
86	38	36.2	119	2 T34696	hypothetical prote
87	38	36.2	142	2 T10078	hypothetical prote
88	38	36.2	152	2 AE2333	hypothetical prote
89	38	36.2	252	1 H69002	conserved hypotet
90	38	36.2	258	2 E69019	conserved hypotet
91	38	36.2	282	2 F64695	biotin synthetase
92	38	36.2	282	2 H71823	lipid A biosynthes
93	38	36.2	290	2 E71631	hypothetical prote
94	38	36.2	297	2 AB3037	nitrate transport
95	38	36.2	299	2 A98249	mannosephosphate i
96	38	36.2	299	2 C69254	probable transposi
97	38	36.2	345	2 A26491	deoxyhypusine synt
98	38	36.2	353	2 T47195	deoxyhypusine synt
99	38	36.2	353	2 T51022	hypothetical prote
100	38	36.2	367	2 T26794	hypothetical prote

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C:Genetics:
A:Gene: map1

Query Match 43.8%; Score 46; DB 2; Length 284;
Best Local Similarity 58.3%; Pred. No. 7.2;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWD 14
| |||:||||
Db 63 TRAVFLKKDWD 74

RESULT 7
B95284
probable [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pSymA
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: B95284
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: B95284
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK64836.1; PID:g14523249; GSPDB:GN001165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hymann, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMA0329
A:Superfamily: plasmid
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 42.9%; Score 45; DB 2; Length 253;
Best Local Similarity 60.0%; Pred. No. 9.3;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 10 EODWDRCVIS 19
:|||||:|
Db 103 DQDWDECIAS 112

RESULT 8
GNBY1
genome polyprotein - coxsackievirus B1
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core pro
polymerase (EC 2.7.7.48)
C:Species: coxsackievirus B1
C>Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999
C:Accession: A26353
R:Iizuka, N.; Kuge, S.; Nomoto, A.
Virology 156, 64-73, 1987
A:Title: Complete nucleotide sequence of the genome of coxsackievirus B1.
A:Reference number: A26353; MUID:87122157; PMID:3027969
A:Accession: A26353
A:Molecule type: genomic RNA
A:Residues: 1-2182 <IIZ>
A:Cross-references: GB:M16560; NID:g323417; PIDN:AAC00531.1; PID:g323418
C:Superfamily: poliovirus genome polyprotein

C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase
F:1-69/Product: coat protein 1A #status predicted <PIA>
F:70-332/Product: coat protein 1B #status predicted <PIB>
F:333-570/Product: coat protein 1C #status predicted <PIC>
F:571-848/Product: coat protein 1D #status predicted <PID>
F:849-998/Product: coat protein 2A #status predicted <C2A>
F:999-1097/Product: core protein 2B #status predicted <C2B>
F:1098-1426/Product: core protein 2C #status predicted <C2C>
F:1427-1515/Product: genome-linked protein Vpg #status predicted <P3A>
F:1516-1537/Product: protein 3A #status predicted <C2B>
F:1538-1720/Product: proteinase #status predicted <PNS>
F:1721-2182/Product: RNA-directed RNA polymerase #status predicted <RNS>
F:1518/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

Query Match 42.4%; Score 44.5; DB 1; Length 2182;
Best Local Similarity 28.6%; Pred. No. 1e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 19; Gaps 1;

QY 2 TTGVGVFG-----IEQDWDRCV 17
||| ||
Db 846 TTGAFGQSGAVYGVYVNRVNRHLATREDWQRCV 880

RESULT 9
B90456
hypothetical protein SSO2792 [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: B90456
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redde
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: B90456
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-498 <KUR>
A:Cross-references: GB:AE006641; NID:g13816138; PIDN:AAK42905.1; GSPDB:GN001155
C:Genetics:
A:Gene: SSO2792

Query Match 41.0%; Score 43; DB 2; Length 498;
Best Local Similarity 38.9%; Pred. No. 40;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVGVFGIEQDWDRCVI 18
:| ||:| | :||:
Db 351 DTPGVYVNVVDPNKCVL 368

RESULT 10
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN001145; UWGP
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 20615

Query Match 41.0%; Score 43; DB 2; Length 5188;

Best Local Similarity 50.0%; Pred. No. 4.4e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWDRCV 17
|||::|||
Db 4803 TTSGVAAMDYDWDGAV 4818

RESULT 11

F90696
hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90696
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: F90696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA833965.1; PID:gl3360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs0542

Query Match 41.0%; Score 43; DB 2; Length 5291;

Best Local Similarity 50.0%; Pred. No. 4.5e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWDRCV 17
|||::|||
Db 4906 TTSGVAAMDYDWDGAV 4921

RESULT 12

E84430
probable inositol polyphosphate-5-phosphatase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84430
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; PMID:20083487; PMID:10617197
A:Accession: E84430
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-417 <STO>
A:Cross-references: GB:AE002093; NID:g4522008; PIDN:AAD21781.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g01900
A:Map position: 2

Query Match 40.5%; Score 42.5; DB 2; Length 417;

Best Local Similarity 60.0%; Pred. No. 40;
Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 5 GVFGEQDWDRCVIS 19
|||::|||
Db 144 GINGISQDF-RCIIS 157

RESULT 13

E84885
hypothetical protein At2g45010 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84885

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; PMID:20083487; PMID:10617197
A:Accession: E84885
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <STO>
A:Cross-references: GB:AE002093; NID:g4895249; PIDN:AD32834.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g45010
A:Map position: 2

Query Match 40.0%; Score 42; DB 2; Length 244;

Best Local Similarity 43.8%; Pred. No. 28;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRC 16
|||::|||
Db 61 NWTGTGFCAGDPESC 76

RESULT 14

F83359
hypothetical protein PA2282 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83359
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Kuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; PMID:20437337; PMID:10984043
A:Accession: F83359
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <STO>
A:Cross-references: GB:AB004654; GB:AE004091; NID:g9948311; PIDN:AAG05670.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2282

Query Match 39.0%; Score 41; DB 2; Length 142;

Best Local Similarity 50.0%; Pred. No. 23;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 6 VFGIEQDWDRCV 17
|||::|||
Db 62 VEGLDHWDRLI 73

RESULT 15

AH1882
nitrate transport permease protein nrtB [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH1882
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriqu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; PMID:21595285; PMID:11759840
A:Accession: AH1882
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-279 <KUP>
A:Cross-references: GB:BA000019; PIDN:BA872567.1; PID:g17129955; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:

A:Gene: nrtB
C:Superfamily: Synechococcus nitrate transport protein nrtB

Query Match 39.0%; Score 41; DB 2; Length 279;
Best Local Similarity 69.2%; Pred. No. 47;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDW 13
I:| | | | |
Db 164 NTTGVQIQDY 176

RESULT 16

S68198
probable transcription regulator (carotenoid gene cluster orf 11) - Myxococcus xanthus
C:Species: Myxococcus xanthus
C:Date: 24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 08-Oct-1999
C:Accession: S68198; S67958
R:Botella, J.; Murillo, F.; Ruiz-vazquez, R.
submitted to the EMBL Data Library, November 1994
A:Description: A cluster of structural and regulatory genes for light-induced carotenogenesis
A:Reference number: S68198
A:Accession: S68198
A:Molecule type: DNA
A:Residues: 1-239 <BOT>
A:Cross-references: EMBL:Z21955; NID:g577589; PIDN:CAA79965.1; PID:g577593
R:Botella, J.A.; Murillo, F.J.; Ruiz-Vazquez, R.
Eur. J. Biochem. 233, 238-248, 1995

A:Title: A cluster of structural and regulatory genes for light-induced carotenogenesis
A:Reference number: S67950; MUID:96061955; PMID:7588751
A:Accession: S67958
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 7-68 <BOW>
A:Cross-references: EMBL:Z21955
C:Keywords: DNA binding

Query Match 39.0%; Score 41; DB 2; Length 299;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 VGVFGIEQDWR 15
| | | | |
Db 284 VHVQGEEDWR 295

RESULT 17

AB1656
cystathionine beta-lyase homolog lin1787 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AB1656
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1656
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-390 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC97018.1; PID:g16414274; GSPDB:GN00178
A:Experimental source: strain Clip11262

C:Genetics:
A:Gene: lin1787
C:Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 39.0%; Score 41; DB 2; Length 390;
Best Local Similarity 46.2%; Pred. No. 66;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDW 13
I:| | | | |
Db 228 NSTGGVLGVQDSW 240

RESULT 18

AG1284
cystathionine beta-lyase homolog lmol679 [imported] - Listeria monocytogenes (strain
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AG1284
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1284
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-390 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99757.1; PID:g16411115; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmol679
C:Superfamily: O-succinylhomoserine (thiol)-lyase

Query Match 39.0%; Score 41; DB 2; Length 390;
Best Local Similarity 46.2%; Pred. No. 66;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDW 13
I:| | | | |
Db 228 NSTGGVLGVQDSW 240

RESULT 19

DB7020
probable membrane transport ATPase ML0890 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: DB7020
R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;
A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: AB6909; MUID:21128732; PMID:11234002
A:Accession: DB7020
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-415 <STO>
A:Cross-references: GB:AL450380; NID:g13092962; PIDN:CAC31271.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML0890
C:Superfamily: anion-transporting ATPase arsaI

Query Match 39.0%; Score 41; DB 2; Length 415;
Best Local Similarity 50.0%; Pred. No. 70;
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWDRCVI 18
| | | | |
Db 131 TVGEFATAGRWDRIVV 146

RESULT 20

T07948
mandelonitrile lyase (EC 4.1.2.10) 3 - black cherry

C:Species: Prunus serotina (black cherry)
 C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 17-Mar-2000
 C:Accession: F07948
 R:Hu, Z.; Poulton, J.E.
 Submitted to the EMBL Data Library, July 1997
 A:Description: Prunus serotina (R-+)-mandelonitrile lyase isoform MDL3 precursor, mRNA
 A:Reference number: Z16239
 A:Accession: T07948
 A:Molecule type: mRNA
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Residues: 1-573 <HUZ>
 A:Cross-references: EMBL:AF013161; NID:g2343180; PIDN:AAB67714.1; PID:g2343181
 C:Genetics:
 A:Gene: MDL3
 C:Function:
 A:Description: catalyzes dissociation of (R)-mandelonitrile to hydrogen cyanide and benzaldehyde
 C:Superfamily: alcohol oxidase
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 39.0%; Score 41; DB 2; Length 573;
 Best Local Similarity 50.0%; Pred. No. 98;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWDRCVIS 19
 :||| ||| ||| :|||
 Db 342 STVTVLGITSDFYQCSIS 359

RESULT 21

T08073
 mandelonitrile lyase (EC 4.1.2.10) 2 precursor - black cherry
 C:Species: Prunus serotina (black cherry)
 C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 17-Mar-2000
 C:Accession: T08073
 R:Hu, Z.; Poulton, J.E.
 Submitted to the EMBL Data Library, December 1997
 A:Reference number: Z16335
 A:Accession: T08073
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-576 <HUZ>
 A:Cross-references: EMBL:AF040078; NID:g2773273; PIDN:AAB96763.1; PID:g2773274
 C:Function:
 A:Description: catalyzes the dissociation of (S)-(-)-mandelonitrile to benzaldehyde and
 A:Pathway: hydrolysis of cyanogenic disaccharides
 C:Superfamily: alcohol oxidase
 C:Keywords: aldehyde-lyase; carbon-carbon lyase

Query Match 39.0%; Score 41; DB 2; Length 576;
 Best Local Similarity 50.0%; Pred. No. 98;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWDRCVIS 19
 :||| ||| ||| :|||
 Db 343 STVTVLGITSDFYQCSIS 360

RESULT 22

JC2248
 polyprotein - sweet potato feathery mottle virus (fragment)
 N:Contains: coat protein; nuclear inclusion b protein
 C:Species: sweet potato feathery mottle virus
 C:Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 17-Nov-2000
 C:Accession: JC2248; PC2164
 R:Morl, M.; Usudi, T.; Hayashi, T.; Nishiguchi, M.
 Biosci. Biotechnol. Biochem. 58, 965-967, 1994
 A:Title: Nucleotide sequence at the 3'-terminal region of sweet potato feathery mottle v
 A:Reference number: JC2248; MUID:94289871; PMID:7764983
 A:Accession: JC2248
 A:Molecule type: mRNA
 A:Residues: 1-688 <NORI>
 A:Cross-references: DBJ:D16664; NID:g391906; PIDN:BAA04068.1; PID:d1004586; PID:g473743
 A:Experimental source: clone pMO-2.3

A>Note: the authors translated the codon CAT for residue 170 as Asp, CAT for residue
 A:Accession: PC2164
 A:Molecule type: protein
 A:Residues: 453-462 <MOR2>
 C:Superfamily: tobacco etch virus genome polyprotein
 C:Keywords: coat protein; polyprotein
 F:1-373/Product: nuclear inclusion b protein (fragment) #status predicted <NIB>
 F:374-688/Product: coat protein #status predicted <COP>
 F:373-374/Cleavage site: Gln-Ser (viral proteinase) #status predicted

Query Match 39.0%; Score 41; DB 2; Length 688;
 Best Local Similarity 47.1%; Pred. No. 1.2e+02;
 Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWDRCVIS 19
 :||| ||| ||| :|||
 Db 73 TVGTFKFGQWDRLLTS 89

RESULT 23

A72479
 hypothetical protein APE2473 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: A72479
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aerc
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: A72479
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-205 <KAW>
 A:Cross-references: DBJ:AP000064; NID:g5105945; PIDN:BA081489.1; PID:g5106178
 C:Genetics:
 A:Experimental source: strain K1
 A:Gene: APE2473
 C:Superfamily: Aeropyrum pernix hypothetical protein APE2473

Query Match 38.6%; Score 40.5; DB 2; Length 205;
 Best Local Similarity 66.7%; Pred. No. 41;
 Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 4 VGVFGIEQDWDNR 15
 :||| ||| :|||
 Db 34 VGVFG-NEDWNR 44

RESULT 24

E95357
 probable [imported] - Sinorhizobium meliloti (strain 1021) magaplasamid pSyma
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: E95357
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; E
 ; Kalman, S.; Keating, D.H.; Palm, C.; Beck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, R.
 Proc.Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
 A:Reference number: A95262; MUID:21396509; PMID:11481432
 A:Accession: E95357
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-259 <KUR>
 A:Cross-references: GB:AF006469; PIDN:AAK65423.1; PID:g14523888; GSPDB:GN00165
 A:Experimental source: strain 1021, megaplasamid pSyma
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMA1398

A:Genome: plasmid

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 38.6%; Score 40.5; DB 2; Length 259;

Best Local Similarity 37.5%; Pred. No. 52;

Matches 9; Conservative 2; Mismatches 6; Indels 7; Gaps 1;

QY 1 NTTVGVG-----IEQWDRCV 17

| ||| | | | | |

DB 92 NAGVGLKRGDLLDTEESWDRCL 115

RESULT 25

C97199

activator of 2-hydroxyglutaryl-CoA dehydratase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: C97199

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: C97199

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-319 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80382.1; PID:g15025444; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2428

Query Match 38.6%; Score 40.5; DB 2; Length 319;

Best Local Similarity 50.0%; Pred. No. 65;

Matches 8; Conservative 2; Mismatches 1; Indels 5; Gaps 1;

QY 1 NTTVGVGIEQWDRC 16

| : | : | | | | |

DB 300 NSTIGAFG-----DRC 310

RESULT 26

AE0471

Probable DEAD-box helicase rhlB [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C:Accession: AE0471

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, M.T.G.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AE0471

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-428 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC93337.1; PID:g15981784; GSPDB:GN00175

C:Genetics:

A:Gene: rhlB

Query Match 38.6%; Score 40.5; DB 2; Length 428;

Best Local Similarity 63.6%; Pred. No. 87;

Matches 7; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

QY 9 IEQDW-DRCVI 18

||:|:| |||:|

DB 251 IEEWPDRCII 261

RESULT 27

S27772

vitellogenin precursor - boll weevil

C:Species: Anthonomus grandis (boll weevil)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S27772

R:Trewitt, P.M.; Hellmann, L.J.; Kumaran, A.K.

submitted to the EMBL Data Library, October 1991

A:Description: Nucleotide sequence and structure of the vitellogenin gene from the boll

A:Reference number: S27772

A:Accession: S27772

A:Molecule type: DNA

A:Residues: 1-1790 <TRE>

A:Cross-references: EMBL:M72980; NID:g155707; PIDN:AAA27740.1; PID:g155708

C:Genetics: 11/1; 678/3; 744/2; 845/2; 902/1; 1466/1

A:introns: 11/1; 678/3; 744/2; 845/2; 902/1; 1466/1

C:Superfamily: boll weevil vitellogenin

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-1790/Product: vitellogenin #status predicted <MAT>

Query Match 38.6%; Score 40.5; DB 1; Length 1790;

Best Local Similarity 61.5%; Pred. No. 3.8e+02;

Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 GVFGI-EQWDRC 16

|| | : || | | |

DB 1122 GVAGVWEQLWDKC 1134

RESULT 28

TTAZ1

toxin I - snake-locks sea anemone

C:Species: Anemonia sulcata (snake-locks sea anemone)

C:Date: 31-Oct-1979 #sequence_revision 31-Oct-1979 #text_change 04-Oct-1996

R:Wunderer, G.; Eulitz, M.

Eur. J. Biochem. 89, 11-17, 1978

A:Title: Amino-acid sequence of toxin I from Anemonia sulcata.

A:Reference number: A01796; MUID:79024590; PMID:29753

A:Accession: A01796

A:Molecule type: protein

A:Residues: 1-46 <WUN>

A:Note: 3-Pro was found in approximately 20% of the molecules

C:Superfamily: sea anemone toxin

C:Keywords: venom

F:4-43,6-34,27-44/Disulfide bonds: #status predicted

Query Match 38.1%; Score 40; DB 1; Length 46;

Best Local Similarity 42.9%; Pred. No. 11;

Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 TVGVFGIEQWDRC 16

| : | | | | |

DB 21 TIWVFGCPSGWNNC 34

RESULT 29

AD2997

hypothetical protein Atu3582 [imported] - Agrobacterium tumefaciens (strain C58, Dupo

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002

C:Accession: AD2997

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McC.

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kar

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AD2997

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-139 <KUR>

A:Cross-references: GB:AE008689; PIDN:AAL44394.1; PID:g17741992; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu3582

A:Map position: linear chromosome

Query Match 38.1%; Score 40; DB 2; Length 139;

Best Local Similarity 60.0%; Pred. No. 33;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 VGVFGIEQDW 13

||:| |:||

Db 9 VGIQIDMDW 18

RESULT 30

E98286

hypothetical protein AGR_L_2491 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002

C:Accession: E98286

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: E98286

A:Molecule type: DNA

A:Residues: 1-141 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK89815.1; PID:g15159745; GSPDB:GN00170

C:Genetics:

A:Gene: AGR_L_2491

A:Map position: linear chromosome

Query Match 38.1%; Score 40; DB 2; Length 141;

Best Local Similarity 60.0%; Pred. No. 34;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 VGVFGIEQDW 13

||:| |:||

Db 11 VGIQIDMDW 20

RESULT 31

FEYTA

ferredoxin [3Fe-4S][4Fe-4S], zinc-containing [validated] - Thermoplasma acidophilum

C:Species: Thermoplasma acidophilum

C:Date: 19-Feb-1984 #sequence_revision 04-Feb-2000 #text_change 16-Jun-2000

C:Accession: T37333; A00222

R:Casper, N.J.; Stalhandske, C.M.V.; Iwasaki, H.; Oshima, T.; Scott, R.A.; Iwasaki, T.

J. Biol. Chem. 274, 23160-23168, 1999

A:Title: Structural conservation of the isolated zinc site in archaeal zinc-containing f

A:Reference number: Z21695; MUID:99367440; PMID:10438486

A:Accession: T37333

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-143 <COS>

A:Cross-references: EMBL:AB023294; NID:g5689050; PIDN:BAAG2797.1; PID:g5689051

A:Experimental source: Strain HO-62

R:Wakabayashi, S.; Fujimoto, N.; Wada, K.; Matsubara, H.; Kersch, L.; Oesterheld, D.

FEBS Lett. 162, 21-24, 1983

A:Title: Amino acid sequence of a ferredoxin from thermoacidophilic archaeobacteria, Ther

A:Reference number: A00222

A:Accession: A00222

A:Molecule type: protein

A:Residues: 2-101, 'Q', 103-105, 'E', 107-143 <KAW>

A:Experimental source: strain DSM 1728

A:Note: the authors believe this ferredoxin has two 4Fe-4S clusters

R:Iwasaki, T.; Suzuki, T.; Kon, T.; Imai, T.; Urushiyama, A.; Ohmori, D.; Oshima, T.

J. Biol. Chem. 272, 3453-3458, 1997

A:Title: Novel zinc-containing ferredoxin family in thermoacidophilic archaea.

A:Reference number: A59164; MUID:97166191; PMID:9013590

A:Contents: annotation; metal binding sites

A:Note: the protein is shown to have one zinc, one 3Fe-4S cluster and one 4Fe-4S clus

C:Comment: For the structure of a closely related sequence with [3Fe-4S][4Fe-4S] clus

C:Genetics:

A:Gene: zfx

C:Superfamily: Sulfolobus zinc-containing ferredoxin; ferredoxin 2[4Fe-4S] homology

C:Keywords: 3Fe-4S; 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein; z

F:2-143/Product: ferredoxin [3Fe-4S][4Fe-4S] #status experimental <MAT>

F:62-142/Domains: ferredoxin 2[4Fe-4S] homology <FER>

F:31,34,58,117/Binding site: zinc (His, His, Asp) #status experimental

F:69,75,134/Binding site: 3Fe-4S cluster (Cys) (covalent) #status experimental

F:79,124,127,130/Binding site: 4Fe-4S cluster (Cys) (covalent) #status experimental

Query Match 38.1%; Score 40; DB 1; Length 143;

Best Local Similarity 38.9%; Pred. No. 34;

Matches 7; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 2 TTVGVFG--IEQDMDRCV 17

||:| |:|||

Db 53 TKLGIHGHVAVDMDCCI 70

RESULT 32

H95187

ribulose-phosphate 3-epimerase family protein [imported] - Streptococcus pneumoniae (

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001

C:Accession: H95187

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzappl

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris

A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: H95187

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-231 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75697.1; PID:g14973105; GSPDB:GN00164; TIGR:

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SPI616

C:Superfamily: yeast ribulose-5-phosphate-epimerase

Query Match 38.1%; Score 40; DB 2; Length 231;

Best Local Similarity 43.8%; Pred. No. 56;

Matches 7; Conservative 3; Mismatches 2; Indels 4; Gaps 1;

QY 5 GVGFGIEQD---WDRC 16

||:| |:|||

Db 201 GLFGLDDDIKAWDIC 216

RESULT 33

A10406

Probable dehydrogenase YPO3351 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001

C:Accession: A10406

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G

li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: A10406

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-256 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC92581.1; PID:g15981278; GSPDB:GN00175

Search completed: October 6, 2003, 07:49:39
Job time : 14.209 secs

C:Genetics:
A:Gene: YPO3351

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 38.1%; Score 40; DB 2; Length 256;

Best Local Similarity 75.0%; Pred. NO. 62;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 10 EQDWDRCV 17

|||||:

Db 111 EQDWDRTI 118

RESULT 34

A90030

hypothetical protein SA2103 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: A90030

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A99758; MUID:21311952; PMID:11418146

A:Accession: A90030

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-315 <KUR>

A:Cross-references: GB:BA000018; PID:g13702110; PIDN:BA043402.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA2103

C:Superfamily: Bacillus subtilis probable transcription regulator yvhJ

Query Match 38.1%; Score 40; DB 2; Length 315;

Best Local Similarity 46.2%; Pred. NO. 77;

Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWDR 15

I:||||:|

Db 70 TIALFGVDSADR 82

RESULT 35

F70357

lipoprotein - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999

C:Accession: F70357

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: F70357

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-349 <AQF>

A:Cross-references: GB:AE000700; NID:g2983248; PIDN:AAC06844.1; PID:g2983249; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: nlpDI

Query Match 38.1%; Score 40; DB 2; Length 349;

Best Local Similarity 43.8%; Pred. NO. 86;

Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWDRCVI 18

I:||||:|

Db 312 TIGYVGIRPDEGRAL 327

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:02 ; Search time 5.38806 Seconds
(without alignments)
165.831 Million cell updates/sec

Title: US-09-765-739A-5

Perfect score: 105

Sequence: 1 NTTGVFGIEQDWDRCVIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	47	44.8	1 MENG_DEIRA	Q9rw10 deinococcus
2	44.5	42.4	1 POUX_CXB1J	P08291 c genome po
3	42	40.0	1 SORC_SCHJA	Q94743 schistosoma
4	41	39.0	1 DPV4_MOUSE	Q35098 mus musculus
5	41	39.0	1 MDL3_PRUSE	P52707 prunus sero
6	41	39.0	1 MDL2_PRUSE	O50048 prunus sero
7	40.5	38.6	1 VIT_ANTGR	Q05808 anthonomus
8	40	38.1	1 TXAL_ANESU	P01533 anemonia su
9	40	38.1	1 FER_THEAC	P00218 thermoplasm
10	40	38.1	1 VIPL_HUMAN	Q9h0v9 homo sapien
11	40	38.1	1 GCST_BACTN	Q89vz6 bacteroides
12	40	38.1	1 FLGE_TREPH	Q56326 treponema p
13	40	38.1	1 HXTD_YEAST	P42833 saccharomyc
14	40	38.1	1 DPV4_RAT	Q62951 rattus norv
15	40	38.1	1 M2B1_MOUSE	O09159 mus musculus
16	39	37.1	1 NRTB_PHOLA	Q51881 phormidium
17	39	37.1	1 MLTB_ECOLI	P41052 escherichia
18	39	37.1	1 Y028_ARCFU	O30207 archaeoglob
19	39	37.1	1 YA67_METTH	O27139 methanobact
20	39	37.1	1 P2X3_HUMAN	P56373 homo sapien
21	39	37.1	1 P2X3_RAT	P49654 rattus norv
22	39	37.1	1 SYFB_AERPE	Q9v9i3 aeropyrum p
23	39	37.1	1 YLX5_CAEEL	P46501 caenorhabdi
24	39	37.1	1 R24L_ARATH	Q9c646 arabidopsis
25	39	37.1	1 RDL2_ARATH	Q8w3k3 arabidopsis
26	39	37.1	1 AFSR_STRCO	P25941 streptomyce
27	39	37.1	1 S190_YEAST	P36123 saccharomyc
28	38.5	36.7	1 FLCA_AQUAE	O67005 aquifex aeo
29	38.5	36.7	1 SYR_AERPE	Q9yb39 aeropyrum p
30	38	36.2	1 B10B_HELPJ	Q9zjk8 helicobacte
31	38	36.2	1 B10B_HELPJ	O25956 helicobacte
32	38	36.2	1 DHYS_NEUCR	P49365 neurospora
33	38	36.2	1 PUR2_RALSO	Q8xxc4 ralstonia s

34	38	36.2	1 Y588_METJA	Q58008 methanococ
35	38	36.2	1 V51K_ACLSP	P27739 apple chlor
36	38	36.2	1 EX5A_ECOLI	P04993 escherichia
37	38	36.2	1 UVRB_TREPA	O83154 treponema p
38	38	36.2	1 FHUE_ECOLI	P16869 escherichia
39	38	36.2	1 OSTA_XYLFA	Q9pf41 xyliella fas
40	38	36.2	1 CLOC_DROME	O61735 drosophila
41	38	36.2	1 EMB5_CAEEL	P34703 caenorhabdi
42	38	36.2	1 RRPL_TOSV	P37800 toscana vir
43	38	36.2	1 PKDR_MOUSE	Q920t6 mus musculu
44	37.5	35.7	1 NO20_MEDTR	P93329 medicago tr
45	37.5	35.7	1 RH23_SCHPO	O74803 schizosacch
46	37.5	35.7	1 CHSD_PETHY	P22925 petunia hyb
47	37.5	35.7	1 HSI_HUMAN	P14317 homo sapien
48	37.5	35.7	1 CHDB_HUMAN	Q9hck8 homo sapien
49	37.5	35.7	1 POLG_CXB5P	Q03053 c genome po
50	37.5	35.7	1 POLG_CXB20	Q9y1g5 c genome po
51	37	35.2	1 CBBY_RHOSH	P95649 rhodobacter
52	37	35.2	1 YQEU_BACSU	P54461 bacillus su
53	37	35.2	1 DHB8_MOUSE	P50171 mus musculu
54	37	35.2	1 PORI_RHOBL	P39767 rhodopseudo
55	37	35.2	1 VP19_HSV6U	P52348 human herpe
56	37	35.2	1 DHYS_METJA	Q58224 methanococ
57	37	35.2	1 RFNG_CHICK	O12972 gallus gall
58	37	35.2	1 P2X4_HUMAN	Q99571 homo sapien
59	37	35.2	1 DHYS_DROME	Q9v5f4 drosophila
60	37	35.2	1 VE2_PAPVE	P11329 european el
61	37	35.2	1 YF02_MYCPN	P75285 mycoplasma
62	37	35.2	1 YB01_MYCPN	P75568 mycoplasma
63	37	35.2	1 YVFO_BACSU	O07013 bacillus su
64	37	35.2	1 YC05_MYCPN	P75571 mycoplasma
65	37	35.2	1 YHJA_ECOLI	P37197 escherichia
66	37	35.2	1 MTHC_DROME	P83119 drosophila
67	37	35.2	1 HSF3_ARATH	O81821 arabidopsis
68	37	35.2	1 IF37_MOUSE	O70194 mus musculu
69	37	35.2	1 IF37_HUMAN	O15371 homo sapien
70	37	35.2	1 CHOD_BREST	P22637 brevibacter
71	37	35.2	1 POT1_SCHPO	O13988 schizosacch
72	37	35.2	1 DPV4_HUMAN	O14531 homo sapien
73	37	35.2	1 SYA_BORBU	O51238 borrelia bu
74	37	35.2	1 V001_VARV	P34010 variola vir
75	37	35.2	1 SYGB_VIBCH	Q9kw8 vibrio chol
76	37	35.2	1 PURL_ZYMMO	Q9req6 zymomonas m
77	37	35.2	1 PURL_AGR5	Q8ueb0 agrobacteri
78	37	35.2	1 OSTA_XANCP	Q8pce0 xanthomonas
79	37	35.2	1 OSTA_XANAC	Q8p222 xanthomonas
80	37	35.2	1 N1RB_KLEPN	Q06458 klebsiella
81	37	35.2	1 ST31_HUMAN	Q9bxui homo sapien
82	37	35.2	1 BXCN_CIOBO	P46081 clostridium
83	37	35.2	1 RPOC_MYCPN	P75271 mycoplasma
84	37	35.2	1 IGIR_HUMAN	P08069 homo sapien
85	37	35.2	1 IGIR_RAT	P24062 rattus norv
86	37	35.2	1 IGIR_MOUSE	Q60751 mus musculu
87	37	35.2	1 RPOC_CHLPN	Q92999 chlamydia p
88	37	35.2	1 POLN_LORDV	P54634 lordsdale v
89	37	35.2	1 YLJ2_CAEEL	P34367 caenorhabdi
90	36.5	34.8	1 RF3_PSEAE	Q9hxb0 pseudomonas
91	36.5	34.8	1 YELH_SCHPO	O13881 schizosacch
92	36.5	34.8	1 SRCB_MOUSE	Q60598 mus musculu
93	36	34.3	1 HSTC_YEREN	O50319 yersinia en
94	36	34.3	1 GVM1_HALNL	P24377 halobacteri
95	36	34.3	1 GVPN_HALME	Q02238 halobacteri
96	36	34.3	1 TAT_VILV1	P23428 visna lenti
97	36	34.3	1 TAT_VILV2	P23429 visna lenti
98	36	34.3	1 FOUB_ECOLI	P31055 escherichia
99	36	34.3	1 UL92_HCMVA	P16798 human cytom
100	36	34.3	1 PIGF_MOUSE	Q09101 mus musculu

ALIGNMENTS

```
MENG_DEIRA
ID MENG_DEIRA STANDARD; PRT; 160 AA.
AC Q9R10;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable S-adenosylmethionine:2-demethylmenaquinone methyltransferase
DE (EC 2.1.1.-).
GN MENG OR DR0859.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RA White O., Eisen J.A., Heidelberger J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Swinn W.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Makarewicz J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.,
RA "Genome sequence of the radioresistant bacterium Deinococcus
RA radiodurans R1", 1577(1999).
RL Science 286:1571-1577(1999).
CC -1- FUNCTION: CONVERTS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK)
CC (by similarity).
CC -1- PATHWAY: Menaquinone biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE MENG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001940; AAF10437.1; -
CC PIR; A75466; A75466.
CC TIGR; DR0859; -.
CC HAMAP; MF_00471; -.
CC InterPro; IPR005493; Methyltransf_6.
CC Pfam; PF03737; Methyltransf_6; 1.
CC Menaquinone biosynthesis; Transferase; Methyltransferase;
CC Complete proteome.
CC SEQUENCE 160 AA; 16860 MW; 54D6F226CA19EC0E CRC64;
CC -----
Query Match 44.8%; Score 47; DB 1; Length 160;
Best Local Similarity 46.7%; Pred. No. 1.2;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 4 VGVFGIEQDWRCVI 18
DB 79 LGVFGVNGWEGVII 93
RESULT 2
ID POLG_CXB1J STANDARD; PRT; 2182 AA.
AC P08291;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Picornain 2A
DE (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein
DE P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28)
DE (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
DE Coxsackievirus B1 (strain Strand).
OC Viruses; ssRNA positive strand viruses, no DNA stage; Picornaviridae;
```

```
FT CHAIN 999 1097 CORE PROTEIN P2B.
FT CHAIN 1098 1426 CORE PROTEIN P2C.
FT CHAIN 1427 1515 CORE PROTEIN P3A.
FT CHAIN 1516 1537 GENOME-LINKED PROTEIN VPG.
FT CHAIN 1538 1720 PICORNAIN 3C. RNA POLYMERASE.
FT CHAIN 1721 2182 RNA-DIRECTED RNA POLYMERASE.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1684 1684 PROTEASE 3C (POTENTIAL).
FT ACT_SITE 1698 1698 PROTEASE 3C (POTENTIAL).
SQ SEQUENCE 2182 AA; 243945 MW; 387B9391275859B1 CRC64;

Query Match 42.4%; Score 44.5; DB 1; Length 2182;
Best Local Similarity 28.6%; Pred. No. 45;
Matches 10; Conservative 1; Mismatches 5; Indels 19; Gaps 1;

QY 2 TTGVGFG-----IEQDMDRCV 17
|| || ||
Db 846 TTGAFGQSGAVYGVNVRVNRHLATREDMORCV 880
|| || ||

RESULT 3
SORC_SCHJA STANDARD; PRT; 171 AA.
ID SORC_SCHJA
AC Q94743;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sorcin.
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Chinese;
RA Clough K.A., Brindley P.J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS, BELONGS
CC TO THE SORCIN/GRANCALIN/CALPAIN LIGHT CHAIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39069; AAB17908.1; -.
DR HSSP; F30626; 1JUO.
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; ehand; 4.
DR ProDom; PD000012; EF-hand; 2.
DR SMART; SM00054; EFh; 4.
DR PROSITE; PS00018; EF_HAND; 3.
KW Calcium-binding; Repeat.
FT CA_BIND 16 27 EF-HAND 1 (POTENTIAL).
FT CA_BIND 53 64 EF-HAND 2 (POTENTIAL).
FT CA_BIND 83 94 EF-HAND 3 (POTENTIAL).
SQ SEQUENCE 171 AA; 19687 MW; 251FB948FA07ADDD CRC64;

Query Match 40.0%; Score 42; DB 1; Length 171;
Best Local Similarity 43.8%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 NTFGVFGIEQDMDRC 16
| : | : | | |
Db 63 NEFLGLFKYVQDWQTC 78

RESULT 4
DPY4_MOUSE STANDARD; PRT; 572 AA.
ID DPY4_MOUSE
```

```
AC O35098; O08886;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Dihydropyrimidinase related protein-4 (DRP-4) (ULIP4 protein).
GN DPYSL4 OR ULIP4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Hamajima N., Kato Y., Kouwaki M., Wada Y., Sasasaki M., Nonaka M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=98314496; PubMed=96523388;
RA BYK T., Ozon S., Sobel A.;
RT "The Ulip family phosphoproteins -- common and specific properties.";
RL Eur. J. Biochem. 254:14-24(1998).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB006715; BAA21888.1; -.
DR EMBL; Y09079; CAA70299.1; -.
DR MGD; MGI:1349764; Dpysl4.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR005847; Urease.
DR Pfam; PF01579; Amidohydro_1; 1.
DR ProDom; PD000518; Urease; 1.
FT CONFLICT 125 126 ER -> DG (IN REF. 2).
FT CONFLICT 354 354 G -> V (IN REF. 2).
FT CONFLICT 420 420 F -> I (IN REF. 2).
SQ SEQUENCE 572 AA; 61961 MW; 37671129FC02C7AF CRC64;

Query Match 39.0%; Score 41; DB 1; Length 572;
Best Local Similarity 47.4%; Pred. No. 44;
Matches 9; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 5 GVFGIEQD---WDRCVIS 19
|| ||| :|||
Db 354 GVGIERMSVWVKCVAS 372

RESULT 5
MDL3 PRUSE STANDARD; PRT; 573 AA.
AC P52707;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE (R)-mandelonitrile lyase isoform 3 precursor (EC 4.1.2.10)
DE (Hydroxynitrile lyase 3) ((R)-oxynitrilase 3).
GN MDL3.
OS Prunus serotina (Black cherry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX NCBI_TaxID=23207;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98076189; PubMed=9414550;
RA Hu Z., Poulton J.E.;
```

RT Sequencing, genomic organization, and preliminary promoter analysis
 RL of a black cherry (R)-(-)-mandelonitrile lyase gene.";
 CC Plant Physiol. 115:1359-1369(1997).
 CC
 CC -1- FUNCTION: Involved in cyanogenesis, the release of HCN from injured
 CC tissues. Catalyzes the stereospecific addition of HCN to a variety
 CC of aldehydes in vitro. It is a major seed constituent, and could
 CC have the additional role of a storage form for reduced nitrogen.
 CC -1- CATALYTIC ACTIVITY: Mandelonitrile + cyanide + benzaldehyde.
 CC (by similarity).
 CC COFACTOR: FAD.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: PRIMARILY FOUND WITHIN PROTEIN BODIES OF THE
 CC COTYLEDONARY PARENCHYMA CELLS WITH LESSER AMOUNTS WITHIN THE
 CC PROCAMBIUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U51562; AAA96782.1; -
 CC EMBL; AF013161; AAB67714.1; -
 CC PIR; T07948; T07948.
 CC InterPro: IPR000172; GMC_oxred.
 CC Pfam; PF05199; GMC_oxred_C; 1.
 CC Pfam; PF00732; GMC_oxred_N; 1.
 CC PROSITE; PS00623; GMC_OXRED.1; 1.
 CC PROSITE; PS00624; GMC_OXRED.2; 1.
 CC Lysase; Glycoprotein; FAD; Flavoprotein; Signal; Multigene family.
 CC SIGNAL 1 27 POTENTIAL.
 CC FT CHAIN 28 573 (R)-MANDELONITRILE LYASE ISOFORM 3.
 CC FT NP_BIND 55 92 FAD (ADP PART) (PROBABLE).
 CC FT CARBOHYD 70 70 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 75 75 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 145 145 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 150 150 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 162 162 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 218 218 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 252 252 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 267 267 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 309 309 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 380 380 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 467 467 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SQ SEQUENCE 573 AA; 62179 MW; DB181C68FED3F800 CRC64;
 CC
 CC Query Match 39.0%; Score 41; DB 1; Length 573;
 CC Best Local Similarity 50.0%; Pred. No. 44;
 CC Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 CC
 CC QY 2 TTGVFGIEQDWDRCVIS 19
 CC Db 342 STVTVLGITSDFYQCSIS 359
 CC
 CC RESULT 6
 CC MDL2_PROSE STANDARD; PRT; 576 AA.
 CC AC O50048;
 CC DT 15-DEC-1998 (Rel. 37, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE (R)-mandelonitrile lyase isoform 2 precursor (EC 4.1.2.10)
 CC GN (Hydroxynitrile lyase 2) ((R)-oxynitrilase 2).
 CC OS Prunus serotina (Black cherry).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.

OX NCBI_TaxID=23207;
 RN [1]
 RA SEQUENCE FROM N.A.
 RP Hu 2; Poulton J.E.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Involved in cyanogenesis, the release of HCN from injured
 CC tissues. Catalyzes the stereospecific addition of HCN to a variety
 CC of aldehydes in vitro. It is a major seed constituent, and could
 CC have the additional role of a storage form for reduced nitrogen
 CC (by similarity).
 CC -1- CATALYTIC ACTIVITY: Mandelonitrile + cyanide + benzaldehyde.
 CC (by similarity).
 CC COFACTOR: FAD.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: PRIMARILY FOUND WITHIN PROTEIN BODIES OF THE
 CC COTYLEDONARY PARENCHYMA CELLS WITH LESSER AMOUNTS WITHIN THE
 CC PROCAMBIUM (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF040078; AAB96763.1; -
 CC EMBL; AF040079; AAB96764.1; -
 CC PIR; T08073; T08073.
 CC InterPro: IPR000172; GMC_oxred.
 CC Pfam; PF05199; GMC_oxred_C; 1.
 CC Pfam; PF00732; GMC_oxred_N; 1.
 CC PROSITE; PS00623; GMC_OXRED.1; 1.
 CC PROSITE; PS00624; GMC_OXRED.2; 1.
 CC Lysase; Glycoprotein; FAD; Flavoprotein; Signal; Multigene family.
 CC SIGNAL 1 28 POTENTIAL.
 CC FT CHAIN 29 576 (R)-MANDELONITRILE LYASE ISOFORM 2.
 CC FT NP_BIND 56 83 FAD (ADP PART) (PROBABLE).
 CC FT CARBOHYD 31 31 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 76 76 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 146 146 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 219 219 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 268 268 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 310 310 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 421 421 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 468 468 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SQ SEQUENCE 576 AA; 62724 MW; 570DC7853AE4D3EA CRC64;
 CC
 CC Query Match 39.0%; Score 41; DB 1; Length 576;
 CC Best Local Similarity 50.0%; Pred. No. 44;
 CC Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 CC
 CC QY 2 TTGVFGIEQDWDRCVIS 19
 CC Db 343 STVTVLGITSDFYQCSIS 360
 CC
 CC RESULT 7
 CC VIT_ANTGR STANDARD; PRT; 1790 AA.
 CC AC Q05808;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Vitellogenin precursor [Contains: YP47; YP160].
 CC GN VTG.
 CC OS Anthonomus grandis (Boll weevil).

```
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
OC Phyltophaga; Curculionidae; Curculioninae; Anthonomini; Anthonomus.
OX NCBI_TaxID=7044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92277665; PubMed=1593641;
RA Trewhitt P.M., Heilmann L.J., Degruillier S.S., Kumaran A.K.;
RT "The boll weevil vitellogenin gene: nucleotide sequence, structure,
RT and evolutionary relationship to nematode and vertebrate vitellogenin
RT genes.";
RL J. Mol. Evol. 34:478-492(1992).
CC -!- FUNCTION: PRECURSOR OF THE EGG-YOLK PROTEINS THAT ARE SOURCES OF
CC NUTRIENTS DURING EMBRYONIC DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Contains 1 VWFD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M72980; AAA27740.1; -.
DR PIR; S27772; S27772.
DR InterPro; IPR001747; Lipid_transprt_N.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF01347; Vitellogenin_N; 1.
DR Pfam; PF00094; vwd; 1.
DR SMART; SM00638; LPD_N; 1.
DR SMART; SM00216; VWD; 1.
DR Glycoprotein; Storage protein; Signal.
KW GLYCOPROTEIN; STORAGE PROTEIN; SIGNAL.
FT SIGNAL 1 19
FT CHAIN 20 1790
FT CHAIN ? 1790
FT CHAIN ? 1790
FT CARBOHYD 219 219
FT CARBOHYD 297 297
FT CARBOHYD 554 554
FT CARBOHYD 573 573
FT CARBOHYD 893 893
FT CARBOHYD 1345 1345
FT CARBOHYD 1416 1416
FT CARBOHYD 1430 1430
FT CARBOHYD 1480 1480
FT CARBOHYD 1699 1699
FT CARBOHYD 1735 1735
FT CARBOHYD 306 309
FT DOMAIN 381 385
FT DOMAIN 1039 1042
FT DOMAIN 1790 AA; 205858 MW; E0B07A5B1E70D0DA CRC64;
SQ SEQUENCE 1790 AA; 205858 MW; E0B07A5B1E70D0DA CRC64;

Query Match 38.6%; Score 40.5; DB 1; Length 1790;
Best Local Similarity 61.5%; Pred. No. 1.7e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 5 GVFGI-EQDWDRG 16
|||:|||||:|
Db 1122 GVAGWELQWDRK 1134

RESULT 8
TXAL_ANESU STANDARD; PRT; 46 AA.
ID TXAL_ANESU STANDARD; PRT; 46 AA.
AC P01533;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurotoxin 1 (Toxin ATX-I).
OS Anemonia sulcata (Snake-locks sea anemone).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actiniaria;
```

```
OC Nynanthaeae; Actiniidae; Anemonia.
OX NCBI_TaxID=6108;
RN [1]
RP SEQUENCE.
RC TISSUE=Cnidoblast;
RX MEDLINE=79024590; PubMed=29753;
RA Wunderer G., Eullitz M.;
RT "Amino-acid sequence of toxin I from Anemonia sulcata.";
RL Eur. J. Biochem. 89:11-17(1978).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=88111667; PubMed=2892680;
RA Widmer H., Wagner G., Schweitz H., Lazdunski M., Wuehrlich K.;
RT "The secondary structure of the toxin ATX Ia from Anemonia sulcata in
RT aqueous solution determined on the basis of complete
RT sequence-specific 1H-NMR assignments.";
RL Eur. J. Biochem. 171:177-192(1988).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=90160299; PubMed=2576133;
RA Widmer H., Billeter M., Wuehrlich K.;
RT "Three-dimensional structure of the neurotoxin ATX Ia from Anemonia
RT sulcata in aqueous solution determined by nuclear magnetic resonance
RT spectroscopy.";
RL Proteins 6:357-371(1989).
CC -!- FUNCTION: BINDS SPECIFICALLY TO THE SODIUM CHANNEL, THEREBY
CC DELAYING ITS INACTIVATION DURING SIGNAL TRANSDUCTION. THUS
CC IT STRONGLY STIMULATES MAMMALIAN CARDIAC MUSCLE CONTRACTION.
CC -!- SUBCELLULAR LOCATION: Secreted; chidocyst.
CC -!- SIMILARITY: BELONGS TO THE SEA ANEMONE SODIUM CHANNEL INHIBITORY
CC TOXIN FAMILY.
CC PIR; A01796; TZA21.
CC PDB; 1ATX; 15-JUL-91.
CC InterPro; IPR000693; Anemone_toxin.
CC Pfam; PF00706; toxin_4; 1.
KW Toxin; Neurotoxin; Ionic channel inhibitor; Sodium channel inhibitor;
KW Chidocyst; 3D-structure.
FT DISULFID 4 43
FT DISULFID 6 34
FT DISULFID 27 44
FT VARIANT 3 3 A -> P (ABOUT 20% OF THE MOLECULES).
FT STRAND 3 5
FT TURN 7 8
FT TURN 14 15
FT STRAND 18 23
FT TURN 29 30
FT STRAND 32 34
FT STRAND 42 46
SQ SEQUENCE 46 AA; 4814 MW; 862C1E21FDA5432D CRC64;

Query Match 38.1%; Score 40; DB 1; Length 46;
Best Local Similarity 42.9%; Pred. No. 4.9;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDWDRG 16
|:|:|:|:|
Db 21 TIWVFGCPGWNNC 34

RESULT 9
FER_THEAC STANDARD; PRT; 142 AA.
ID FER_THEAC STANDARD; PRT; 142 AA.
AC P00218; Q9V304;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE zinc-containing ferredoxin.
GN ZFX OR TA1448.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
```


RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerk A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May function as a regulator of ERGIC-53.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein; endoplasmic
CC reticulum.
CC -!- SIMILARITY: Contains 1 legumainous lectin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ549957; CAD71268.1; -;
CC EMBL: ALI36617; CAB6552.1; -;
CC EMBL: AK075347; BAC11559.1; -;
CC EMBL: BC000347; AAH00347.1; ALT_INIT.
CC EMBL: BC005822; AAH05822.1; ALT_INIT.
CC EMBL: BC005862; AAH05862.1; ALT_INIT.
CC InterPro: IPR005052; Lectin_leg.
CC Pfam: PF03388; Lectin_leg-like; 1.
CC Signal: Glycoprotein; Lectin; Transmembrane; Endoplasmic reticulum.
KW SIGNAL 1 44
FT CHAIN 45 348
FT DOMAIN 45 313
FT DOMAIN 314 336
FT DOMAIN 337 348
FT DOMAIN 49 276
FT SITE 344 346
FT SITE 200 237
FT DISULFID 181 181
FT CARBOHYD 344 346
FT MUTAGEN 4 4
FT CONFLICT 4 4
FT CONFLICT 288 288
FT CONFLICT 348 AA; 39711 MW; F79382E20AE67751 CRC64;
SQ

Query Match 38.1%; Score 40; DB 1; Length 348;
Best Local Similarity 46.7%; Pred. No. 39;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 TVGVFGIEQDNDRCV 17
Db 4 TLGPLGSQQWRRCL 18

RESULT 11
GCST_BACTN
ID GCST_BACTN STANDARD; PRT; 361 AA.
AC Q89Y26;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable aminomethyltransferase (EC 2.1.2.10) (Glycine cleavage system
DE T protein).
DE GCVT OR B74584.
GN Bacteroides thetaiotaomicron.
OS Bacterioidetes; Bacteroides (class); Bacteroidales;
OC Bacterioidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).

CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of
CC glycine (By similarity).
CC -!- CATALYTIC ACTIVITY: (6S)-tetrahydrofolate + S-
CC aminomethylidihydrofolate + dihydrodipolypyrrole.
CC methylenetetrahydrofolate + NH(3) + dihydrodipolypyrrole.
CC -!- SUBUNIT: The glycine cleavage system is composed of four proteins:
CC P, T, L, and H (By similarity).
CC -!- SIMILARITY: Belongs to the gcvt family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE016945; AAO79689.1; -;
CC HAMAP: MF_00259; -; 1.
CC TRANSFERASE; Aminotransferase; Complete proteome.
SQ SEQUENCE 361 AA; 39825 MW; C57E87A164F95A89 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 361;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 VGVFGIEQDNDRCV 17
Db 111 VNAANIERDNNVCV 124

RESULT 12
FLGE_TREPH STANDARD; PRT; 463 AA.
AC Q56326;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Flagellar hook protein flgE.
GN FLGE.
OS Treponema phagedenis.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=162;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kazan 5;
RX MEDLINE=94266716; PubMed=8206841;
RA Limberger R.J., Slivinski L.L., Samsonoff W.A.;
RT "Genetic and biochemical analysis of the flagellar hook of Treponema
RT phagedenis.";
RL J. Bacteriol. 176:3631-3637(1994).
CC -!- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U04619; AAA73467.1; -;
CC InterPro: IPR001444; Flag_bb_rod.
CC Pfam: PF00460; flg_bb_rod; 1.
CC PROSITE: PS00588; FLAGELLIA_BB_ROD; 1.
KW Flagella.
SQ SEQUENCE 463 AA; 49527 MW; 960548C364CA35BC CRC64;

Query Match 38.1%; Score 40; DB 1; Length 463;
Best Local Similarity 54.5%; Pred. No. 51;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
QY      2  TTGVFGIEQD 12
      11 | | | | |
Db      113  TTAGAFGVDRD 123

RESULT 13
HXTD_YEAST      STANDARD;      PRT;      540 AA.
AC      P42833;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Hexose transporter HXT14.
GN      HXT14 OR HXT9 OR YNI218C OR N0345.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288c / FY1676;
RA      MEDLINE=96076632; PubMed=7502583;
RX      Maftahi M., Nicaud J.-M., Levesque H., Gaillardin C.;
RT      "Sequencing analysis of a 24.7 kb fragment of yeast chromosome XIV
RT      identifies six known genes, a new member of the hexose transporter
RT      family and ten new open reading frames.";
RL      Yeast 11:1077-1085(1995).
RC      CC -!- FUNCTION: PROBABLE GLUCOSE TRANSPORTER.
CC      CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC      CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC      CC -----
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      CC the European Bioinformatics Institute. There are no restrictions on its
CC      CC use by non-profit institutions as long as its content is in no way
CC      CC modified and this statement is not removed. Usage by and for commercial
CC      CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      CC or send an email to license@isb-sib.ch)
CC      CC -----
DR      EMBL; Z46259; -; NOT_ANNOTATED_CDS.
DR      EMBL; Z71595; CAA96250.1; -
DR      PIR; S63299; S63299.
DR      SGD; S0005262; HXT14.
DR      GO; GO:0005354; F:galactose transporter activity; IGI.
DR      GO; GO:0008645; F:hexose transport; IGI.
DR      InterPro; IPR007114; MFS.
DR      InterPro; IPR005828; Sub-transporter.
DR      InterPro; IPR005829; Sug-transporter.
DR      Pfam; PF00083; Sugar_tr; 1.
DR      PRINTS; PR00171; SUGTRNSPORT.
DR      TIGRAME; TIGR000879; SP; 1.
DR      PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR      PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
DR      Repeat; Transmembrane; Sugar transport; Transport.
FT      DOMAIN      1 56
FT      TRANSMEM      57 76
FT      DOMAIN      77 119
FT      TRANSMEM      120 140
FT      DOMAIN      141 146
FT      TRANSMEM      147 167
FT      DOMAIN      168 177
FT      TRANSMEM      178 198
FT      DOMAIN      199 204
FT      TRANSMEM      205 225
FT      DOMAIN      226 243
FT      TRANSMEM      244 264
FT      DOMAIN      265 357
FT      TRANSMEM      358 370
FT      DOMAIN      375 380
FT      TRANSMEM      381 398
FT      DOMAIN      399 405
FT      TRANSMEM      406 426

QY      2  TTGVFGIEQD 12
      11 | | | | |
Db      113  TTAGAFGVDRD 123

RESULT 14
DPY4_RAT
ID      DPY4_RAT      STANDARD;      PRT;      564 AA.
AC      Q62951;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      15-JUL-1998 (Rel. 36, Last annotation update)
DE      Dihydropyrimidinase related protein-4 (DRP-4) (Collapsin response
DE      mediator protein 3) (CRMP-3) (Fragment).
GN      DPYSL4.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RX      MEDLINE=96424532; PubMed=8815901;
RT      Wang M., Strittmatter S.M.;
RT      "A family of rat, CRMP genes is differentially expressed in the
RT      nervous system."
RT      J. Neurosci. 16:6197-6207(1996).
CC      CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC      CC -!- DEVELOPMENTAL STAGE: EXPRESSED TRANSIENTLY IN DEVELOPING SPINAL
CC      CC CORD AND SELECTIVELY IN THE POSTNATAL CEREBELLUM.
CC      CC -!- SIMILARITY: BELONGS TO THE DEHYDROPYRIMIDINASE FAMILY.
CC      CC -----
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      CC the European Bioinformatics Institute. There are no restrictions on its
CC      CC use by non-profit institutions as long as its content is in no way
CC      CC modified and this statement is not removed. Usage by and for commercial
CC      CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      CC or send an email to license@isb-sib.ch)
CC      CC -----
DR      EMBL; U52103; AAB03281.1; -
DR      InterPro; IPR006680; Amidohydro_1.
DR      InterPro; IPR005847; Urease.
DR      Pfam; PF01979; Amidohydro_1; 1.
DR      ProDom; PD000518; Urease; 1.
FT      NON_TER      1
FT      TRANSMEM      564 AA; 61085 MW; 42050891CC1436D2 CRC64;
SQ      SEQUENCE

Query Match      38.1%; Score 40; DB 1; Length 540;
Best Local Similarity 42.1%; Pred. No. 63;
Matches 8; Conservative 4; Mismatches 3; Indels 4; Gaps 1;

QY      5  GVGFGIEQD---WDRCVIS 19
      11 | | | | |
Db      346  GINGIERMSVWVEKCVAS 364

RESULT 15
M2BI_MOUSE
ID      M2BI_MOUSE      STANDARD;      PRT;      1013 AA.
```

AC O09159; O55037; Q64443; Q9DBQ1;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lysosomal alpha-mannosidase precursor (EC 3.2.1.24) (Mannosidase, alpha
DE alpha B) (lysosomal acid alpha-mannosidase) (Laman) (Mannosidase, alpha
DE class 2B member 1).
GN MAN2B1 OR MAN2B OR MANB OR LAMAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Macrophage;
RX MEDLINE=97454441; PubMed=9355733;
RA Beccari T., Appolloni M.G., Costanzi E., Stinchi S., Stirling J.L.,
RA Della Fazio M.A., Servillo G., Viola M.P., Orlacchio A.;
RT "Lysosomal alpha-mannosidases of mouse tissues: characteristics of the
RT isoenzymes, and cloning and expression of a full-length cDNA.";
RL Biochem. J. 327:45-49(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Stinchi S., Stirling J.L., Orlacchio A., Beccari T.;
RT "Promoter characterization and structure of the gene encoding mouse
RT lysosomal alpha-mannosidase.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE OF 3-1013 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97449314; PubMed=9305783;
RA Merkle R.K., Zhang Y., Ruest P.J., Lal A., Liao Y.-F., Moremen K.W.;
RT "Cloning, expression, purification, and characterization of the murine
RT lysosomal acid alpha-mannosidase.";
RL Biochim. Biophys. Acta 1336:132-146(1997).
CC -!- FUNCTION: NECESSARY FOR THE CATABOLISM OF N-LINKED CARBOHYDRATES
CC RELEASED DURING GLYCOPROTEIN TURNOVER.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC mannose residues in alpha-D-mannosides.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U29947; AAC53369.1; ALT_INIT.
DR EMBL; AF044192; AAC78560.1; .
DR EMBL; AF044174; AAC78560.1; JOINED.
DR EMBL; AF044175; AAC78560.1; JOINED.
DR EMBL; AF044176; AAC78560.1; JOINED.
DR EMBL; AF044177; AAC78560.1; JOINED.
DR EMBL; AF044178; AAC78560.1; JOINED.
DR EMBL; AF044179; AAC78560.1; JOINED.
DR EMBL; AF044180; AAC78560.1; JOINED.
DR EMBL; AF044181; AAC78560.1; JOINED.
DR EMBL; AF044182; AAC78560.1; JOINED.
DR EMBL; AF044183; AAC78560.1; JOINED.
DR EMBL; AF044184; AAC78560.1; JOINED.
DR EMBL; AF044185; AAC78560.1; JOINED.
DR EMBL; AF044186; AAC78560.1; JOINED.
DR EMBL; AF044187; AAC78560.1; JOINED.
DR EMBL; AF044188; AAC78560.1; JOINED.
DR EMBL; AF044189; AAC78560.1; JOINED.
DR EMBL; AF044190; AAC78560.1; JOINED.
DR EMBL; AF044191; AAC78560.1; JOINED.
DR EMBL; AK004817; BAB23588.1; .
DR EMBL; U87240; AAC09470.1; ALT_INIT.
DR PIR; T42385; T42385.
DR MGI; MGI:107286; Man2bl.
DR InterPro; IPR000602; Glyco_hydro_38.
DR Pfam; PF01074; Glyco_hydro_38; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Lysosome; Zymogen; Signal.
FT SIGNAL 1 43 POTENTIAL.
FT CHAIN 44 1013 LYSOSOMAL ALPHA-MANNOSIDASE.
FT CARBOHYD 310 310 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 693 693 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 767 767 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 931 931 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 45 49 PGARA -> LASG (IN REF. 3).
FT CONFLICT 344 344 A -> Y (IN REF. 3).
FT CONFLICT 980 980 F -> Y (IN REF. 3).
SQ SEQUENCE 1013 AA; 114604 MW; CE8243B103A4F64 CRC64;
Query Match 38.1%; Score 40; DB 1; Length 1013;
Best Local Similarity 47.1%; Pred. No. 1.le+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 3 TVGVFGIEQDWDRCVIS 19
||| : || : |||
DB 716 TVGPIPRDDWGKEVIS 732
RESULT 16
NRTB_PHOLA
ID NRTB_PHOLA STANDARD; PRT; 279 AA.
AC Q51881;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nitrate transport permease protein nrtb.
GN NRTB.
OS Phormidium laminosum.
OC Bacteria; Cyanobacteria; Oscillatoriales; Phormidium.
OX NCBI_TaxID=32059;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OH-1-P-CL1;
RX MEDLINE=95375238; PubMed=7647306;
RA Merchan F., Kindle K.L., Llama M.J., Serra J.L., Fernandez E.;
RT "Cloning and sequencing of the nitrate transport system from the

SEQ	SEQUENCE	361 AA;	40256 MW;	7E98F040504F75F9	CRC64;
Query Match		37.1%;	Score 39;	DB 1;	Length 361;
Best Local Similarity		50.0%;	Pred. No. 58;		
Matches		6;	Conservative	2;	Mismatches 4; Indels 0; Gaps 0;
QY	4	GVGFIEQDWDR	15		
		: : :			
DB	156	VGLIGVETRWGR	167		
RESULT 18					
Y028_ARCFU		STANDARD;	PRT;	371 AA.	
ID	Y028_ARCFU				
AC	O30207;				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Hypothetical protein AF0028.				
DN	AF0028.				
OS	Archaeoglobus fulgidus.				
OC	Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;				
OC	Archaeoglobaceae; Archaeoglobus.				
NCBI_Taxid=2234;					
OX	[1]				
RN	SEQUENCE FROM N.A.				
RP	STRAIN=VC-16 / DSM 4304 / ATCC 49558;				
RC	MEDLINE=98049343; PubMed=9389475;				
RX	Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,				
RA	Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,				
RA	Richardson D.L., Kurlavage A.R., Graham D.E., Kyriakides N.C.,				
RA	Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,				
RA	Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,				
RA	Peterson S., Reich C.T., McNeil L.K., Badger J.H., Glodek A., Zhou L.,				
RA	Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,				
RA	Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,				
RA	Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,				
RA	Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,				
RA	Venter J.C.;				
RT	"The complete genome sequence of the hyperthermophilic, sulphate-				
RT	reducing archaeon Archaeoglobus fulgidus.";				
RL	Nature 390:364-370(1997).				
CC	-!- SIMILARITY: STRONG, TO M.JANNASCHII MJ1678, M.THERMOAUTOTROPHICUM				
CC	MTH1067 AND A.FULGIDUS AF0181.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; AE001105; AAB91200.1; -.				
DR	PIR; D69253; D69253.				
DR	TIGR; AF0028; -.				
KW	Hypothetical protein; Complete proteome.				
SEQ	SEQUENCE 371 AA; 41930 MW; B22056D00437BEAA				CRC64;
Query Match		37.1%;	Score 39;	DB 1;	Length 371;
Best Local Similarity		33.3%;	Pred. No. 60;		
Matches		5;	Conservative	6;	Mismatches 4; Indels 0; Gaps 0;
QY	4	GVGFIEQDWDR	CVI 18		
		: : : :			
DB	217	IGVMGMKEEAEIVL	231		
RESULT 19					
YA67_METTH		STANDARD;	PRT;	377 AA.	
ID	YA67_METTH				
AC	O27139;				
DT	15-JUL-1998 (Rel. 36, Created)				

RL Nature 377:432-435(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Dorsal root ganglion;
 RX MEDLINE=96008344; PubMed=7566119;
 RA Chen C.-C., Akopian A.N., Sivillotti L., Colquhoun D., Burnstock G.,
 Wood J.N.;
 RT "A P2X purinoceptor expressed by a subset of sensory neurons.";
 RL Nature 377:428-432(1995).
 CC -!- FUNCTION: RECEPTOR FOR ATP THAT ACTS AS A LIGAND GATED ION CHANNEL.
 CC -!- SUBUNIT: HOMO- OR HETEROPOLYMERS.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN SENSORY GANGLIA.
 CC -!- SIMILARITY: BELONGS TO THE P2X RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X91167; CAA62594.1; -;
 CC EMBL; X90651; CAA62223.1; -;
 CC PIR; I58099; I58099.
 CC InterPro: IPR001429; P2X_receptor.
 CC Pfam; PF00864; P2X_receptor; 1.
 CC PRINTS; PR01307; P2XRECEPTOR.
 CC TIGRFAMs; TIGR00863; P2X; 1.
 CC PROSITE; PS01212; P2X_RECEPTOR; 1.
 CC Ionic channel; Transmembrane; Ion transport; Receptor; Glycoprotein.
 CC DOMAIN 1 22 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 23 43 1 (POTENTIAL).
 CC DOMAIN 44 317 EXTRACELLULAR, CYSTEINE-RICH (POTENTIAL).
 CC TRANSMEM 318 338 2 (POTENTIAL).
 CC DOMAIN 339 397 CYTOPLASMIC (POTENTIAL).
 CC FT CARBOHYD 139 139 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 194 194 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CONFLICT 96 96 M -> I (IN REF. 2).
 CC SEQUENCE 397 AA; 44392 MW; AA7ED5410F639D25 CRC64;
 CC -----
 CC Query Match 37.1%; Score 39; DB 1; Length 397;
 CC Best Local Similarity 34.8%; Pred. No. 64;
 CC Matches 8; Conservative 4; Mismatches 3; Indels 8; Gaps 1;
 CC -----
 CC QY 3 TVGVFGI-----EQWDRCV 17
 CC | | | | : : | : | :
 CC 235 TGGVLGKIGVWCDDLKAWDQCI 257
 CC -----
 CC RESULT 22
 CC SYFB_AERPE
 CC ID SYFB_AERPE STANDARD; PRT; 548 AA.
 CC AC Q9Y9I3;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Phenylalanyl-L-TRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
 CC trNA ligase beta chain) (PheRS).
 CC GN PHE1 OR APE2305
 CC OS Aeropyrum pernix.
 CC OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
 CC OC Desulfurococccaceae; Aeropyrum.
 CC OC NCBI_TaxID=56636;
 CC [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=KI;
 CC RX MEDLINE=99310339; PubMed=10382966;
 CC Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,

RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo H. Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
 CC -!- diphosphate + L-phenylalanyl-tRNA(Phe).
 CC -!- SUBUNIT: tetramer of two alpha and two beta chains (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
 CC CHAIN FAMILY. SUBFAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AP000064; BAA81317.1; -;
 CC PIR; E72457; E72457.
 CC HAMAP; MF_00284; -; 1.
 CC InterPro: IPR005147; B5.
 CC InterPro: IPR004531; PheT_arch.
 CC Pfam; PF03484; B5; 2.
 CC TIGRFAMs; TIGR00471; pheT_arch; 1.
 CC KW Aminoacyl-L-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 CC Complete proteome.
 CC SEQUENCE 548 AA; 60100 MW; 8884A2CEF49937C4 CRC64;
 CC -----
 CC Query Match 37.1%; Score 39; DB 1; Length 548;
 CC Best Local Similarity 50.0%; Pred. No. 89;
 CC Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC -----
 CC QY 5 GVFGTEQWD 14
 CC | : | : | : | :
 CC 64 GIAGVEEGWD 73
 CC -----
 CC RESULT 23
 CC YLX5_CAEEL
 CC ID YLX5_CAEEL STANDARD; PRT; 751 AA.
 CC AC P46501;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
 CC DT 01-NOV-1995 (Rel. 32, Last annotation update)
 CC DE Hypothetical 84.8 kDa protein F23F12.5 in chromosome III.
 CC GN F23F12.5.
 CC OS Caenorhabditis elegans.
 CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC OC Rhabditidae; Peloderinae; Caenorhabditis.
 CC OX NCBI_TaxID=6239;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=Bristol N2;
 CC Du Z.;
 CC Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U12965; AAA20607.1; -;

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana";
RL Nature 408:816-820(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=98191999; PubMed=9530866;
RA Aarts M.G., te Lintel Hekkert B., Holub E.B., Beynon J.L.,
RA Stiekema W.J., Pereira A.;
RT "Identification of R-gene homologous DNA fragments genetically linked
RT to disease resistance loci in Arabidopsis thaliana";
RL Mol. Plant Microbe Interact. 11:251-258(1998).
CC -!- FUNCTION: Potential disease resistance protein.
CC -!- DOMAIN: The LRR repeats probably act as specificity determinant of
CC pathogen recognition (by similarity).
CC -!- SIMILARITY: Belongs to the disease resistance NB-LRR family.
CC -!- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC -!- SIMILARITY: Contains 1 NB-ARC domain.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -!- DATABASE: NAME=NIB-LRRS;
CC NOTE=Functional and comparative genomics of disease resistance gene
CC homologs;
CC WWW="http://niblrrs.ucdavis.edu".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC082643; AAG50637.1; ALT_SEQ.
DR EMBL; AB077822; BAB38572.1; -.
DR EMBL; AF039380; AAC14556.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR002182; NB-ARC.
DR Pfam; PF00560; LRR; 1.
DR Pfam; PF00931; NB-ARC; 1.
DR Hypothetical protein; Plant defense; ATP-binding; Repeat;
KW Leucine-rich repeat.
FT DOMAIN 10 45 LEUCINE-ZIPPER.
FT REPEAT 148 460 NB-ARC.
FT REPEAT 580 604 LRR 1.
FT REPEAT 605 628 LRR 2.
FT REPEAT 847 872 LRR 3.
FT NP_BIND 191 198 ATP (POTENTIAL).
FT CONFLICT 265 265 E -> G (IN REF. 3).
FT CONFLICT 307 307 A -> T (IN REF. 3).
FT CONFLICT 320 320 E -> G (IN REF. 3).
FT SEQUENCE 910 AA; 105677 MW; 91A6E59ADFABBD3 CRC64;

Query Match 37.1%; Score 39; DB 1; Length 910;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 7 FGIEDWDR 15
DB 8 FGVEKLWDR 16
RESULT 26
AFSR_STRCO
ID AFSR_STRCO STANDARD; PRT; 993 AA.
AC P25941; P08467; Q9L000;
DT 01-AUG-1988 (Rel. 08, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Regulatory protein afsR.
GN AFSR OR AFSB OR SCO4426 OR SCD6.04C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=91071605; PubMed=2253887;
RA Horinouchi S., Kito M., Nishiyama M., Furuya K., Hong S.-K.,
RA Miyake K., Beppu T.;
RT "Primary structure of AfsR, a global regulatory protein for secondary
RT metabolite formation in Streptomyces coelicolor A3(2).";
RL Gene 95:49-56(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
RN [3]
RP SEQUENCE OF 751-993 FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=87008389; PubMed=2428809;
RA Horinouchi S., Suzuki H., Beppu T.;
RT "Nucleotide sequence of afsR, a pleiotropic gene involved in
RT secondary metabolism in Streptomyces coelicolor A3(2) and
RT 'Streptomyces lividans'";
RL J. Bacteriol. 168:257-269(1986).
CC -!- FUNCTION: GLOBAL REGULATORY PROTEIN FOR SECONDARY METABOLITE
CC FORMATION.
CC -!- PTM: PHOSPHORYLATED BY AFSK.
CC -!- MISCELLANEOUS: THE N-TERMINAL AND THE C-TERMINAL REGIONS ARE
CC INDEPENDENTLY CAPABLE OF DIRECTING ACTINORHODIN PRODUCTION.
CC -!- SIMILARITY: BELONGS TO THE AFSR/DNRI/REDD FAMILY OF REGULATORS.
CC -!- SIMILARITY: Contains 3 TPR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90155; BAA14186.1; -.

```
DR EMBL; AL939120; CAB88433.1; -.
DR EMBL; M16011; AAA26694.1; -.
DR PIR; A25037; A25037.
DR PIR; J00488; J00488.
DR PhosSite; P25941; -.
DR InterPro; IPR005158; BAD.
DR InterPro; IPR000767; Disease_resist.
DR InterPro; IPR002182; NB-ARC.
DR InterPro; IPR001440; TPR.
DR Pfam; PF03704; BTAD; 1.
DR Pfam; PF00931; NB-ARC; 1.
DR Pfam; PF00515; TPR; 3.
DR Pfam; PF00486; trans_reg_C; 1.
DR PRINTS; PR00364; DISEASERSIST.
DR SMART; SM00028; TPR; 3.
DR Transcription; regulation; ATP-binding; DNA-binding; Activator; Repeat;
KW Antibiotic biosynthesis; Pigment; Phosphorylation; TPR repeat; Repeat;
KW Complete proteome.
FT REPEAT 833 856 TPR 1.
FT REPEAT 863 896 TPR 2.
FT REPEAT 942 975 TPR 3.
FT NF_BIND 330 337 ATP (POTENTIAL).
FT DNA_BIND 736 811 H-T-H MOTIF (BY SIMILARITY).
FT DNA_BIND 974 988 H-T-H MOTIF (BY SIMILARITY).
FT CONFLICT 256 256 Q -> R (IN REF. 1).
FT CONFLICT 318 318 A -> P (IN REF. 1).
FT CONFLICT 751 766 DVLSVSGRIHAEAEA -> MCSWCPVASSMPRRRH (IN REF. 3).
SQ SEQUENCE 993 AA; 105714 MW; 8793CF27CEA7254C CRC64;

Query Match 37.1%; Score 39; DB 1; Length 993;
Best Local Similarity 50.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTVGVFGIEQDWDRCV 17
Db 641 TAAGVVAIERPGDLV 656

RESULT 27
SI90_YEAST
ID SI90_YEAST STANDARD; PRT; 1033 AA.
AC P36723.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Slt4-associating protein SAPI90.
GN SAPI90 OR YKR028W.
OC Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RP MEDLINE-96220458; PubMed-8649382;
RX Luke M.M., della Seta F., di Como C.J., Sugimoto H., Kobayashi R.,
RA Arndt K.T.;
RT "The SAP, a new family of proteins, associate and function positively
with the Slt4 phosphatase."
RL Mol. Cell. Biol. 16:2744-2755(1996).
[2]
SEQUENCE FROM N.A.
RP Urrestarazu L.A., Jauniaux J.-C.;
RA Submitted (Mar-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ASSOCIATES WITH THE Slt4 PHOSPHATASE IN A CELL CYCLE
DEPENDENT MANNER. MAY BE DIRECTLY OR INDIRECTLY INVOLVED IN
Slt4 DEPENDENT FUNCTIONS IN BUDDING AND IN NORMAL G1 CYCLIN
EXPRESSION.
CC -!- PTM: HYPERPHOSPHORYLATED IN THE ABSENCE OF Slt-4.
CC -!- SIMILARITY: BELONGS TO THE SAPS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U50561; AAC49304.1; -.
DR EMBL; Z28253; CAA82100.1; -.
DR PIR; S38100; S38100.
DR SGD; S0001736; SAPI90.
DR GO; GO:0005737; C:cytoplasm; IPI.
DR GO; GO:0004722; F:protein serine/threonine phosphatase activity; IPI.
DR GO; GO:0000082; P:C1/S transition of mitotic cell cycle; IGI.
DR Pfam; PF00499; SAPS; 1.
KW Phosphorylation; Cell cycle.
FT CONFLICT 1009 1033 DEEDNDMGNEGYSLYRSKEAF -> MKMKRMIVAMRR
ATHYIQEKKSEKLSKHSCHIFTLYSSIFYCSTFAFRSP
IIHQYTFLEACNLILTEYNLIKNEYSLLNLKAFI (IN REF. 2).
SQ SEQUENCE 1033 AA; 117097 MW; 33E5575B58B73B1A CRC64;

Query Match 37.1%; Score 39; DB 1; Length 1033;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCV 18
Db 762 NTVLGDFGNDQESDDVI 779

RESULT 28
FLGA_AQUAE
ID FLGA_AQUAE STANDARD; PRT; 223 AA.
AC Q67005.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagella basal body P-ring formation protein flga precursor.
GN FLGA OR AQ_833.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
[1]
SEQUENCE FROM N.A.
RP STRAIN-VF5;
RX MEDLINE-98196666; PubMed-9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: INVOLVED IN THE ASSEMBLY PROCESS OF THE P-RING
FORMATION. IT MAY ASSOCIATE WITH FLGF ON THE ROD CONSTITUTING A
STRUCTURE ESSENTIAL FOR THE P-RING ASSEMBLY OR MAY ACT AS A
MODULATOR PROTEIN FOR THE P-RING ASSEMBLY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE FLGA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A5000709; AAC06962.1; -.
DR PIR; B70372; B70372.
DR InterPro; IPR004924; Flga.
DR Pfam; PF03240; Flga; 1.
KW Flagella; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1
FT POTENTIAL.
```

```

FT CHAIN ? 223 FLAGELLA BASAL BODY P-RING FORMATION
FT PROTEIN FLGA.
SQ SEQUENCE 223 AA; 26303 MW; 98C0A5189D8254CD CRC64;

Query Match 36.7%; Score 38.5; DB 1; Length 223;
Best Local Similarity 43.8%; Pred. No. 43;
Matches 7; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 NTTGVFGIEQDWDRC 16
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 87 NTKVGTIALELEW-KC 101

RESULT 29
SYR_AERPE
ID SYR_AERPE STANDARD; PRT; 644 AA.
AC Q9YB39;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginyl-tRNA synthetase (EC 6.1.1.19) (Arginine--tRNA ligase) (ArgRS).
GN ARG5 OR APE156.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococccaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1.
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-arginine + tRNA(Arg) = AMP +
CC diphosphate + L-arginyl-tRNA(Arg).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AP000062; BAA80759.1; -.
DR PIR; B72559; B72559.
DR HAMAP; MF_00123; -.
DR InterPro; IPR001278; Arg_trna-synt_1c.
DR InterPro; IPR005148; N.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF03485; N-Arg; 1.
DR Pfam; PF00750; tRNA-synt_ld; 1.
DR PRINTS; PR01038; TRNASYNTHARG.
DR TIGRFAMS; TIGR00456; argS; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 129 139 "HIGH" REGION.
SQ SEQUENCE 644 AA; 72172 MW; 076E18F55C0B5A08 CRC64;

Query Match 36.7%; Score 38.5; DB 1; Length 644;
Best Local Similarity 53.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 3 TVGVFGIE-QDWD 14
```

```

Db 295 TLGREGVDFDWD 307
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 295 TLGREGVDFDWD 307

RESULT 30
BIOB_HELPJ
ID BIOB_HELPJ STANDARD; PRT; 282 AA.
AC Q9ZJK8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).
GN BIOB OR JHP1298.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
CC -!- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -!- PATHWAY: Biotin biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AE001553; AAD06876.1; -.
DR PIR; H71823; H71823.
DR InterPro; IPR002684; Biotin_synth.
DR InterPro; IPR006638; Elp3.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMS; TIGR00433; bioB; 1.
KW Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
FT METAL 17 17 IRON-SULFUR (POTENTIAL).
FT METAL 21 21 IRON-SULFUR (POTENTIAL).
FT METAL 24 24 IRON-SULFUR (POTENTIAL).
SQ SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAF1B CRC64;

Query Match 36.2%; Score 38; DB 1; Length 282;
Best Local Similarity 40.0%; Pred. No. 66;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVFGIEQDWD 14
   11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 157 GIFGLNESWE 166

RESULT 31
BIOB_HELPJ
ID BIOB_HELPJ STANDARD; PRT; 282 AA.
AC Q25956;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).
GN BIOB OR HPI406.
OS Helicobacter pylori (Campylobacter pylori).
```

CC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 CC Helicobacteraceae; Helicobacter.
 CC [Uni_TaxID=210;
 RN SEQUENCE FROM N.A.
 RC STRAIN=26695; ATCC 700392;
 EX MEDLINE=97394467; PubMed=5252185;
 RA Tomb J.-F., White C., Keriavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.P., Gill S., Dougherty B.A.,
 RA Naissang B., Richardson J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftis B., Richardson J., Dodson R., Khakhria H.G., Glodek A.,
 RA McKenney K., Fitzgibbon J., M. Lee W., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson S.D., Kelley J.M.,
 RA Cotton M.D., Weidman J.M., Fujiki C., Bowman C., Watney L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 RT *pylori*.";
 RL Nature 388:539-547(1997).
 CC -1- CATALYTIC ACTIVITY: Dethiobiotin + sulfur -> biotin.
 CC -1- PATHWAY: Biotin biosynthesis; last step.
 CC -1- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000640; AAD08448.1; -
 DR PIR: F64695; F64695.
 DR TIGR: HPI406; -
 DR InterPro: IPR002684; Biotin_synth.
 DR SMART: PF04055; Radical_SAM; 1.
 DR Pfam: SM00729; Elp3; 1.
 DR TIGRfams: TIGR00433; bioB; 1.
 DR Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
 FT METAL 17 17 IRON-SULFUR (POTENTIAL).
 FT METAL 21 21 IRON-SULFUR (POTENTIAL).
 FT METAL 24 24 IRON-SULFUR (POTENTIAL).
 SQ SEQUENCE 282 AA; 31475 MW; 8EAD8B80DF151EE CRC64;
 Query Match 36.2%; Score 38; DB 1; Length 282;
 Best Local Similarity 40.0%; Pred. No. 66;
 Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 5 GVFGIEQDWD 14
 Db 157 GIGLNESE 166
 I::: : : :
 RESULT 32
 ID DHYS_NEUCR STANDARD; PRT; 353 AA.
 AC P49365; O9P3J1;
 DT 01-DEC-1996 (Rel. 33, Created)
 DT 15-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Deoxyhypusine synthase (EC 2.5.1.46) (DHS).
 GN NaS-1 OR B7F21.30.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 CC [Uni_TaxID=5141;
 RN SEQUENCE FROM N.A.
 RA MEDLINE=9602575; PubMed=7592594;
 RT Tao Y., Chen K.Y.;
 RT "Molecular cloning and functional expression of Neurospora

RT deoxyhypusine synthase cDNA and identification of yeast deoxyhypusine
 RT synthase cDNA.";
 RL J. Biol. Chem. 270:23984-23987(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=74-OR23-1A / FGSC 987;
 RA PubMed=12655011;
 RA Mannhaupt G., Montrone C., Haase D., Mewes H.-W., Aign V.,
 RA Hohnsbeil J.D., Fartmann B., Nyakatura G., Kempken F., Maier J.,
 RA Schulte U.;
 RT "What's in the genome of a filamentous fungus? Analysis of the
 RT Neurospora genome sequence.";
 RL Nucleic Acids Res. 31:1944-1954(2003).
 RN [3]
 RP SEQUENCE OF 109-126; 151-171; 217-236 AND 290-305, AND
 RP CHARACTERIZATION.
 RA MEDLINE=95113855; PubMed=7814398;
 RA Tao Y., Chen K.Y.;
 RT Purification of deoxyhypusine synthase from *Neurospora crassa* to
 RT homogeneity by substrate elution affinity chromatography.";
 RL J. Biol. Chem. 270:383-386(1995).
 CC -1- FUNCTION: CATALYZES THE NAD-DEPENDENT OXIDATIVE CLEAVAGE OF
 CC SPERMIDINE AND THE SUBSEQUENT TRANSFER OF A SPECIFIC LYSINE RESIDUE
 CC OF THE EIF-5A PRECURSOR PROTEIN TO FORM THE INTERMEDIATE
 CC DEOXYHYPUSINE RESIDUE.
 CC -1- CATALYTIC ACTIVITY: [eIF5A-precursor]-lysine + spermidine ->
 CC [eIF5A-precursor]-deoxyhypusine + propane-1,3-diamine.
 CC -1- PATHWAY: Hypusine biosynthesis; first step.
 CC -1- SUBUNIT: Homotrimer.
 CC -1- SIMILARITY: Belongs to the deoxyhypusine synthase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U22400; AAC49075.1; -
 DR EMBL: AL389901; CAB97475.1; -
 DR PIR: T47195; T47195.
 DR PIR: T51022; T51022.
 DR HSP: P49366; DHS.
 DR InterPro: IPR002773; DS.
 DR Pfam: PF01916; DS; 1.
 DR ProDom: PD007730; DS; 1.
 DR TIGRfams: TIGR00321; dhys; 1.
 DR Hypusine biosynthesis; Transferase; NAD.
 FT CONFLICT 245 245 K -> E (IN REF. 1).
 FT CONFLICT 311 311 A -> S (IN REF. 1).
 SQ SEQUENCE 353 AA; 38942 MW; A290B0D00022DD13 CRC64;
 Query Match 36.2%; Score 38; DB 1; Length 353;
 Best Local Similarity 50.0%; Pred. No. 83;
 Matches 8; Conservative 4; Mismatches 2; Indels 2; Gaps 1;
 QY 2 TTGVFGIEQDWDRCV 17
 Db 115 TTAG--GIEDDFIKCL 128
 I::: : : :
 RESULT 33
 ID PUR2_RALSO STANDARD; PRT; 422 AA.
 AC Q8XGC4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Phosphoribosylamine--glycine ligase (EC 6.3.4.13) (GARS) (Glycinamide
 DE ribonucleotide synthetase) (Phosphoribosylglycinamide synthetase).

```
GN PURD OR RSC2191 OR RS01408.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- CATALYTIC ACTIVITY: ATP + 5-phospho-D-riboseylamine + glycine = ADP
CC + phosphate + N(1)-(5-phospho-D-riboseyl)glycinamide.
CC -!- PATHWAY: De novo purine biosynthesis; second step.
CC -!- SIMILARITY: Belongs to the GARS family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL646068; CAD15898.1; -.
DR HAMAP; MF_00138; -.
DR InterPro; IPR000115; Gars.
DR Pfam; PF01071; GARS_1.
DR Pfam; PF02842; GARS_B; 1.
DR Pfam; PF02843; GARS_C; 1.
DR Pfam; PF02844; GARS_N; 1.
DR TIGRFAMs; TIGR00877; purD; 1.
DR PROSITE; PS00184; GARS; FALSE_NEG.
KW Purine biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 422 AA; 45015 MW; 1248E0A487C7A297 CRC64;
Query Match 36.2%; Score 38; DB 1; Length 422;
Best Local Similarity 75.0%; Pred. No. le+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 8 GIEQDWDR 15
DB 319 GWELDWDR 326
RESULT 34
Y588_METJA
ID Y588_METJA STANDARD; PRT; 430 AA.
AC Q58008;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0024 protein M0588.
GN M0588.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Klenk H.-P., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Clifton H.P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -!- SIMILARITY: Belongs to the UPF0024 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67507; AAB98579.1; -.
DR PIR; D64373; D64373.
DR TIGR; MJ0588; -.
DR InterPro; IPR001656; UPF0024.
DR Pfam; PF01142; UPF0024; 1.
DR TIGRFAMs; TIGR00094; TIGR00094; 1.
DR PROSITE; PS01268; UPF0024; 1.
DR Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 430 AA; 51214 MW; 436D1B752AA75BCC CRC64;
Query Match 36.2%; Score 38; DB 1; Length 430;
Best Local Similarity 63.6%; Pred. No. le+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 4 VGVFGIEQDWD 14
DB 225 VGRFIERDWE 235
RESULT 35
V51K_ACLSPP
ID V51K_ACLSPP STANDARD; PRT; 460 AA.
AC P27739;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 50.8 kDa protein (ORF2).
OS Apple chlorotic leaf spot virus (isolate plum P863) (ACLSV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Trichovirus.
OX NCBI_TaxID=73473;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=91021011; PubMed=2219716;
RA German S., Candresse T., Lanneau M., Huet J.-C., Pernollet J.-C.,
RA Dunez J.;
RT "Nucleotide sequence and genomic organization of apple chlorotic leaf
RT spot closterovirus.";
RL Virology 179:104-112(1990).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S35.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M58152; AAA42588.1; -.
DR PIR; B45353; B45353.
DR MEROPS; S35.001; -.
DR InterPro; IPR001815; Capilloptases35.
DR Pfam; PF02103; Peptidase_S35; 1.
DR PRINTS; PR00995; CAPILLOPTASE.
KW Hydrolase; Serine protease.
SQ SEQUENCE 460 AA; 50831 MW; E5677CC3092B2C8C CRC64;
```

Query Match 36.2%; Score 38; DB 1; Length 460;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 4 VGVFGIEQD 12
Db 214 VGAFGEQD 222

Search completed: October 6, 2003, 07:44:12
Job time : 9.38806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 9.67164 Seconds
(without alignments)
178.981 Million cell updates/sec

Title: US-09-765-739A-4

Perfect score: 98
Sequence: 1 NPTVALYGLKQDWNGVSA 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	89.8	278	JE0216	28k surface antige
2	61	62.2	280	JE0217	28k surface antige
3	57	58.2	276	JE0218	28k surface antige
4	55	56.1	284	I40882	major antigenic pr
5	51.5	52.6	553	G90593	subtilisin, serine
6	50	51.0	286	JE0219	28k surface antige
7	48	49.0	439	E82426	phosphoglycerate t
8	47	48.0	341	S73685	hypothetical prote
9	46	46.9	133	JE0221	28k surface antige
10	45	45.9	375	AG0606	hypothetical RNA m
11	45	45.9	461	AC0005	probable membrane
12	44	44.9	477	F87329	hypothetical prote
13	44	44.9	1649	C86822	hypothetical prote
14	43.5	44.4	122	G64558	hypothetical prote
15	43.5	44.4	299	T23932	hypothetical prote
16	43	43.9	463	AC0969	probable purine pe
17	43	43.9	480	T15839	hypothetical prote
18	43	43.9	918	S04255	regulatory protein
19	43	43.9	1156	C87371	TonB-dependent rec
20	42.5	43.4	328	G71838	probable nadh oxid
21	42.5	43.4	1465	S31262	TyB protein - years
22	42.5	43.4	1467	PC1253	TyB protein - years
23	42.5	43.4	1802	S52611	TyB protein - years
24	42.5	43.4	1803	S56894	TyB protein - years
25	42	42.9	190	E95093	hydrolase, haloaci
26	42	42.9	190	B97961	phosphoglycolate p
27	42	42.9	298	S75205	hypothetical prote
28	42	42.9	389	T43979	hypothetical prote
29	42	42.9	412	T09313	immediate-early pr

30	42	42.9	420	2	AH2117	hypothetical prote
31	42	42.9	656	1	S59631	endo-1,4-beta-xyla
32	42	42.9	661	1	S59633	endo-1,4-beta-xyla
33	42	42.9	691	2	B75622	hypothetical prote
34	42	42.9	1023	2	A47296	thiazide-sensitive
35	42	42.9	2894	2	C64474	hypothetical prote
36	41.5	42.3	122	2	A17950	hypothetical prote
37	41.5	42.3	253	2	A12523	site-specific DNA-
38	41	41.8	243	2	A72669	probable glutaredo
39	41	41.8	290	2	S76787	hypothetical prote
40	41	41.8	368	2	AD0937	alanine racemase l
41	41	41.8	373	2	G75073	hypothetical prote
42	41	41.8	412	2	T18760	hypothetical prote
43	41	41.8	459	2	T10307	hypothetical prote
44	41	41.8	544	2	S25101	CTP synthase (EC 6
45	41	41.8	585	2	B75265	probable acid-CoA
46	41	41.8	624	2	A83237	hypothetical prote
47	41	41.8	716	2	A75595	hypothetical prote
48	41	41.8	757	2	C84120	subtilisin-type pr
49	41	41.8	1232	2	T21018	hypothetical prote
50	41	41.8	1587	2	AB2012	hypothetical prote
51	40.5	41.3	89	2	E53226	retrovirus-related
52	40.5	41.3	358	2	B64427	carbamoyl-phosphat
53	40.5	41.3	982	2	S00954	pol polyprotein -
54	40.5	41.3	1356	2	F84486	probable retroelem
55	40	40.8	105	2	F83935	hypothetical prote
56	40	40.8	125	2	H83068	hypothetical prote
57	40	40.8	162	2	F82084	probable 2-demethy
58	40	40.8	297	2	AH3649	iron(III)-transpor
59	40	40.8	311	2	A53808	homeotic protein c
60	40	40.8	343	2	S74438	iron(III) dicitrat
61	40	40.8	363	2	D69502	rod shape-determ
62	40	40.8	389	2	F83538	chitin-binding pro
63	40	40.8	412	2	AF3512	membrane-bound lyt
64	40	40.8	413	2	H83833	hypothetical prote
65	40	40.8	436	2	F97058	Fe-S oxidoreductas
66	40	40.8	452	2	B59096	hypothetical prote
67	40	40.8	595	2	A48878	uric acid/xanthine
68	40	40.8	843	2	H72204	pullulanase - Ther
69	40	40.8	991	2	S43891	dna exoribonucleas
70	40	40.8	1060	2	T30823	bumetanide sensiti
71	39.5	40.3	194	2	T52044	dof zinc finger pr
72	39.5	40.3	726	2	S37964	beta-adaptin homol
73	39	39.8	99	2	G69891	hypothetical prote
74	39	39.8	123	2	H75430	conserved hypothet
75	39	39.8	155	2	A12780	hypothetical prote
76	39	39.8	237	2	B97560	hypothetical prote
77	39	39.8	254	2	F69309	ATP-binding protei
78	39	39.8	278	2	C83502	conserved hypothet
79	39	39.8	286	1	RLPUGG	rRNA N-glycosidase
80	39	39.8	293	2	H90344	thiosulfate sulfur
81	39	39.8	313	2	A65233	FLAT element-bindi
82	39	39.8	329	1	G65210	conserved hypothet
83	39	39.8	353	2	E70100	tryptophan-trRNA l
84	39	39.8	358	2	T34382	hypothetical prote
85	39	39.8	372	2	H84637	hypothetical prote
86	39	39.8	374	2	H71091	hypothetical prote
87	39	39.8	464	2	G71123	hypothetical prote
88	39	39.8	482	2	G75483	probable leucyl am
89	39	39.8	493	2	G72800	gp10 protein - Myc
90	39	39.8	560	2	F81423	L-lactate permease
91	39	39.8	567	2	S76847	phosphoglucosylase
92	39	39.8	580	2	B70668	probable Acyl-CoA
93	39	39.8	590	1	WMSECB	64K capsid assembl
94	39	39.8	643	2	S55610	polyprotein - equi
95	39	39.8	671	2	A38109	autolysin - Entero
96	39	39.8	694	2	AB3396	NADH2 dehydrogenas
97	39	39.8	697	2	JC2365	fused proteinase-c
98	39	39.8	732	2	A75219	translation elonga
99	39	39.8	735	2	G71203	probable translati
100	39	39.8	751	2	A13392	phosphoribosylform

ALIGNMENTS

```
RESULT 1
JE0216
28k surface antigen 3 - Ehrlichia chaffensis
N;Alternate names: MAP1
C;Species: Ehrlichia chaffensis
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C;Accession: JE0216
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0216; MUID:98321180; PMID:9647746
A;Accession: JE0216
A;Molecule type: DNA
A;Residues: 1-278 <RED>
A;Cross-references: GB:AF062761

Query Match 89.8%; Score 88; DB 2; Length 278;
Best Local Similarity 83.3%; Pred. No. 1.2e-06;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWGSA 18
Db 60 NPTVALYGLKQDWEGISS 77

RESULT 2
JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N;Alternate names: MAP1
C;Species: Ehrlichia chaffensis
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C;Accession: JE0217
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0217; MUID:98321180; PMID:9647746
A;Accession: JE0217
A;Molecule type: DNA
A;Residues: 1-280 <RED>
A;Cross-references: GB:AF062761

Query Match 62.2%; Score 61; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.03;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
Db 60 NTTGVFGLKQDWMDG 74

RESULT 3
JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N;Alternate names: MAP1
C;Species: Ehrlichia chaffensis
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C;Accession: JE0218
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0218; MUID:98321180; PMID:9647746
A;Accession: JE0218
A;Molecule type: DNA
A;Residues: 1-276 <RED>
A;Cross-references: GB:AF062761

Query Match 58.2%; Score 57; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.13;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 NPTVALYGLKQDWNG 15
Db 59 NTTGVFGLKQNDWG 73

RESULT 4
140882
major antigenic protein - heartwater rickettsia
C;Species: Cowdria ruminantium (heartwater rickettsia)
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40882; S42827
R;van Vliet, A.H.; Jongejan, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A;Title: Molecular cloning, sequence analysis, and expression of the gene encoding the
A;Reference number: I40882; MUID:94178956; PMID:8132352
A;Accession: I40882
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-284 <RES>
A;Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C;Genetics:
A;Gene: map1

Query Match 56.1%; Score 55; DB 2; Length 284;
Best Local Similarity 64.3%; Pred. No. 0.29;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGV 16
Db 63 TKAIVGLKQDWGV 76

RESULT 5
G90593
subtilisin, serine proteinase [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: G90593
R;Chambers, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: G90593
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-553 <KOR>
A;Cross-references: GB:AL445566; PID:g14090070; PIDN:CACL3828.1; GSPDB:GN00153
C;Genetics:
A;Gene: MYPU_6550
A;Genetic code: SGC3

Query Match 52.6%; Score 51.5; DB 2; Length 553;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 9; Conservative 5; Mismatches 3; Indels 1; Gaps 1;

QY 1 NPTVALYGLKO-DWNGVS 17
Db 229 NERIKYGIKRFNWNGIS 246

RESULT 6
JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N;Alternate names: MAP1
C;Species: Ehrlichia chaffensis
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C;Accession: JE0219
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A;Reference number: JE0219; MUID:98321180; PMID:9647746
A;Accession: JE0219
```


A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <RUR>
A:Cross-references: GB:AL590842; PIDN:CAC88901.1; PID:g15978149; GSPDB:GN00175
C:Genetics:
A:Gene: YP00034
C:Superfamily: hypothetical protein b2882

Query Match 45.9%; Score 45; DB 2; Length 461;
Best Local Similarity 61.5%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDWN 14
II IIII
DB 256 PTPLVGLSPDWN 258

RESULT 12

F87329
hypothetical protein CC0649 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: F87329
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
n, J.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87329
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-477 <STO>
A:Cross-references: GB:AE005673; NID:g13421864; PIDN:AAK22634.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0649

Query Match 44.9%; Score 44; DB 2; Length 477;
Best Local Similarity 57.1%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDWN 15
II IIII
DB 254 PTWGLYNSKWDWRG 267

RESULT 13

C86822
hypothetical protein ydbK [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86822
R:Boletini, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1649 <STO>
A:Cross-references: GB:AE005176; PID:g12724583; PIDN:AAK05677.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ydbK

Query Match 44.9%; Score 44; DB 2; Length 1649;
Best Local Similarity 41.2%; Pred. No. 11e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWN 17
:II :I :I :II
DB 1601 SPTTGTGFKLNQNNNSTS 1617

RESULT 14

G64558
hypothetical protein HP0311 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: G64558
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McK
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467; PMID:9252185
A:Accession: G64558
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-122 <TOM>
A:Cross-references: GB:AE000549; GB:AE000511; NID:g2313403; PIDN:AAD07384.1; PID:g231

Query Match 44.4%; Score 43.5; DB 2; Length 122;
Best Local Similarity 56.2%; Pred. No. 9;
Matches 9; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 1 NPTVALYGLKQDWN 15
I :I IIII II I
DB 55 NQILAFYGLKINDWQG 70

RESULT 15

T23932
hypothetical protein R05D7.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T23932
R:Dobson, R.
Submitted to the EMBL Data Library, October 1996
A:Reference number: Z19819
A:Accession: T23932
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-299 <WIL>
A:Cross-references: EMBL:Z81105; PIDN:CAB03219.1; GSPDB:GN00019; CESP:R05D7.4
A:Experimental source: clone R05D7
C:Genetics:
A:Gene: CESP:R05D7.4
A:Map position: 1
A:Introns: 29/3; 122/1; 162/1; 231/2; 268/3
C:Superfamily: tropinesterase

Query Match 44.4%; Score 43.5; DB 2; Length 299;
Best Local Similarity 47.4%; Pred. No. 22;
Matches 9; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

QY 1 NPTVALYGLKQDWN 16
:I :I IIII II I
DB 46 SPLVIVHGLFGQKQNNNSV 64

RESULT 16

AC0969
probable purine permease STY4046 [imported] - Salmonella enterica subsp. enterica ser
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AC0969
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AC0969
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-463 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD03247.1; PID:gl6504869; GSPDB:GN00176
 A:Gene: STY4046
 C:Superfamily: hypothetical protein b2882

Query Match 43.9%; Score 43; DB 2; Length 463;
 Best Local Similarity 61.5%; Pred. No. 43;
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDWN 14
 II III III
 Db 257 PTPLYGLGIDWN 269

RESULT 17

T15839

hypothetical protein C54G7.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 15-Sep-2000

R:Du, Z.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid C54G7.

A:Reference number: Z18416

A:Accession: T15839

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-480 <DUZ>

A:Cross-references: EMBL:U40410; NID:gl065453; PID:gl065454; PIDN:AAA81391.1; CESP:C54G7

C:Genetics:

A:Gene: CESP:C54G7.2

A:Introns: 27/3; 76/2; 102/2; 185/3; 283/1; 384/3

C:Superfamily: Caenorhabditis elegans hypothetical protein C08F8.4

Query Match 43.9%; Score 43; DB 2; Length 480;
 Best Local Similarity 60.0%; Pred. No. 44;
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWN 17
 III I I III
 Db 351 TVATYLSAVWHGVS 365

RESULT 18

S04255

regulatory protein qa-1S - Neurospora crassa

C:Species: Neurospora crassa

C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 21-Jul-2000

C:Accession: S04255; A23941; E31277

R:Geever, R.F.; Huetet, L.; Baum, J.A.; Tyler, B.M.; Patel, V.B.; Rutledge, B.J.; Case, M.

J. Mol. Biol. 207, 15-34, 1989

A:Title: DNA sequence, organization and regulation of the qa gene cluster of Neurospora

A:Reference number: S04250; MUID:89293848; PMID:2525625

A:Accession: S04255

A:Molecule type: DNA

A:Residues: 1-918 <GIL>

A:Cross-references: EMBL:X14603; NID:g3060; PIDN:CAA32753.1; PID:g295929

R:Huetet, L.; Giles, N.H.

Proc. Natl. Acad. Sci. U.S.A. 83, 3381-3385, 1986

A:Title: The qa repressor gene of Neurospora crassa: wild-type and mutant nucleotide sequence

A:Reference number: A23941; MUID:86205901; PMID:3010294

A:Accession: A23941

A:Molecule type: DNA

A:Residues: 1-904,'R',908-918 <GIL2>

A:Cross-references: GB:M13208; NID:gl68869; PIDN:AAA33612.1; PID:gl68870

A:Note: this sequence has been revised in reference S04250

C:Genetics:

A:Gene: qa-1S

A:Introns: 76/1

C:Superfamily: shikimate dehydrogenase homology
 C:Keywords: DNA binding; repressor; transcription regulation
 F:649-877/Domain: shikimate dehydrogenase homology <SKD>

Query Match 43.9%; Score 43; DB 2; Length 918;
 Best Local Similarity 57.1%; Pred. No. 86;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 ALYGLKQDWN 18
 IIII IIII I
 Db 684 ALYGDWDWIGIRA 697

RESULT 19

C87371

TonB-dependent receptor [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: C87371

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: C87371

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1156 <STO>

A:Cross-references: GB:AE005673; NID:gl3422267; PIDN:AAK22967.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0983

Query Match 43.9%; Score 43; DB 2; Length 1156;
 Best Local Similarity 80.0%; Pred. No. 11e+02;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 LYGLKQDWN 15
 I III IIII
 Db 607 LGGLKGDWNG 616

RESULT 20

G71838

probable nadh oxidoreductase I - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: G71838

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: G71838

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-328 <ARN>

A:Cross-references: GB:AE001545; GB:AE001439; NID:g4155776; PIDN:AAD06772.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: nuof

Query Match 43.4%; Score 42.5; DB 2; Length 328;
 Best Local Similarity 64.3%; Pred. No. 36;
 Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 NPTVALYGLK-ODW 13
 IIIII I I I I
 Db 203 NPTVAFYDSKDQEW 216

RESULT 21

S31262
 TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty4 (fragment)
 C:Species: Saccharomyces cerevisiae
 C>Date: 18-Jun-1993 #sequence_revision 25-Apr-1997 #text_change 26-May-2000
 C:Accession: S31262
 R:Janetzky, B.; Lehle, L.
 J. Biol. Chem. 267, 19798-19805, 1992
 A:Title: Ty4, a new retrotransposon from Saccharomyces cerevisiae, flanked by tau-element
 A:Reference number: S31262; PMID:93015829; PMID:1328182
 A:Accession: S31262
 A:Molecule type: DNA
 A:Residues: 1-1465 <JAN>
 A:Cross-references: EMBL:S46865
 C:Genetics:
 A:Mobile element: retrotransposon Ty4
 C:Superfamily: TyB protein

Query Match 43.4%; Score 42.5; DB 2; Length 1465;
 Best Local Similarity 69.2%; Pred. No. 1.7e+02;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 Qy 5 ALYGLKQ---DWN 14
 Db 1077 ALYGLKQSPKEWN 1089
 ||||| :||
 ||||| :||

RESULT 22
 PC1253
 TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty4 (fragment)
 N:Contains: integrase; proteinase; reverse transcriptase; RNase H
 C:Species: Saccharomyces cerevisiae
 C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 26-May-2000
 C:Accession: PC1253; S27465
 R:Stucka, R.; Schwarzlose, C.; Lochmueller, H.; Haecker, U.; Feldmann, H.
 Gene 122, 119-128, 1992
 A:Title: Molecular analysis of the yeast Ty4 element: Homology with Ty1, copia, and plant
 A:Reference number: JC1482; PMID:93083972; PMID:1333437
 A:Accession: PC1253
 A:Molecule type: DNA
 A:Residues: 1-1467 <STW>
 R:Stucka, R.; Schwarzlose, C.; Lochmueller, H.; Hcker, U.; Feldmann, H.
 submitted to the EMBL data Library, May 1992
 A:Description: Molecular analysis of the yeast Ty4 element: homology with Ty1, copia, and
 A:Reference number: S27465
 A:Accession: S27465
 A:Molecule type: DNA
 A:Residues: 1-1467 <STW>
 A:Cross-references: EMBL:M94164
 A:Gene: Ty4B
 A:Mobile element: retrotransposon Ty4
 C:Superfamily: TyB protein

Query Match 43.4%; Score 42.5; DB 2; Length 1467;
 Best Local Similarity 69.2%; Pred. No. 1.7e+02;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 Qy 5 ALYGLKQ---DWN 14
 Db 1077 ALYGLKQSPKEWN 1089
 ||||| :||
 ||||| :||

RESULT 23
 S52611
 TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty4
 N:Alternate names: protein YHL008w-a
 C:Species: Saccharomyces cerevisiae
 C>Date: 05-May-1995 #sequence_revision 19-Oct-1995 #text_change 23-Mar-2001
 C:Accession: S52611
 R:Favell, T.
 submitted to the EMBL Data Library, June 1994
 A:Description: The sequence of S. cerevisiae cosmid L5018.
 A:Reference number: S46798

A:Accession: S52611
 A:Molecule type: DNA
 A:Residues: 1-1802 <FAV>
 A:Cross-references: EMBL:U11581
 C:Genetics:
 A:Map position: 8L
 A:Mobile element: retrotransposon Ty4
 C:Superfamily: TyB protein

Query Match 43.4%; Score 42.5; DB 2; Length 1802;
 Best Local Similarity 69.2%; Pred. No. 2.1e+02;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 Qy 5 ALYGLKQ---DWN 14
 Db 1414 ALYGLKQSPKEWN 1426
 ||||| :||
 ||||| :||

RESULT 24
 S56894
 TyB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty4.JL
 N:Alternate names: protein J0780; protein YJL113w
 C:Species: Saccharomyces cerevisiae
 C>Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-2000
 C:Accession: S56894
 R:Czilepluch, C.; Kordes, E.; Pujol, A.; Jauniaux, J.C.
 submitted to the Protein Sequence Database, September 1995
 A:Reference number: S56891
 A:Accession: S56894
 A:Molecule type: DNA
 A:Residues: 1-1803 <CZI>
 A:Cross-references: EMBL:Z49389
 C:Genetics:
 A:Map position: 10L
 A:Mobile element: retrotransposon Ty4.JL
 C:Superfamily: TyB protein

Query Match 43.4%; Score 42.5; DB 2; Length 1803;
 Best Local Similarity 69.2%; Pred. No. 2.1e+02;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 Qy 5 ALYGLKQ---DWN 14
 Db 1415 ALYGLKQSPKEWN 1427
 ||||| :||
 ||||| :||

RESULT 25
 E95093
 hydrolase, haloacid dehalogenase-like family [Imported] - Streptococcus pneumoniae (s
 C:Species: Streptococcus pneumoniae
 C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
 C:Accession: E95093
 R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
 on, J.D.; Umavam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
 nson, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
 A:Title: Complete Genome Sequence of a Virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: E95093
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-190 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK74942.1; PID:g14972282; GSPDB:GN00164; TIGR:
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0805

Query Match 42.9%; Score 42; DB 2; Length 190;
 Best Local Similarity 57.1%; Pred. No. 25;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 Qy 3 TVALYGLKQDWN 16

Db 31 TLALYGITQDHDV 44

RESULT 26

B97961

phosphoglycolate phosphatase (EC 3.1.3.18) [Imported] - Streptococcus pneumoniae (strain C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001

C:Accession: B97961

R:Hoskins, J.A.; Albora Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B97961

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-190 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99518.1; PID:gl5458305; GSPDB:GN00174

C:Genetics:

A:Gene: gph

C:Keywords: phosphoric monoester hydrolase

Query Match 42.9%; Score 42; DB 2; Length 190;

Best Local Similarity 57.1%; Pred. No. 25;

Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 TVALYGLKQDNGV 16

I:||||: ||: |

Db 31 TLALYGITQDHDV 44

RESULT 27

S75205

hypothetical protein slr2052 - Synecocystis sp. (strain PCC 6803)

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S75205

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S75205

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-298 <KAN>

A:Cross-references: EMBL:D90503; GB:AB001339; NID:gl652127; PIDN:BAAL7119.1; PID:gl65219

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Synecocystis hypothetical protein slr2052

Query Match 42.9%; Score 42; DB 2; Length 298;

Best Local Similarity 42.1%; Pred. No. 39;

Matches 8; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 2 PTVALYG--LKQDNGVSA 18

||||: ||: |

Db 278 PTVIAYGKRISRSWGVQS 296

RESULT 28

T43979

hypothetical protein U19 [Imported] - human herpesvirus 6

C:Species: human herpesvirus 6

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 28-Jul-2000

C:Accession: T43979; T44165

R:isegawa, Y.; Mukai, T.; Nakano, K.; Kagawa, M.; Chen, J.; Mori, Y.; Sunagawa, T.; Kawa

J. Virol. 73, 8053-8063, 1999

A:Title: Comparison of the complete DNA sequences of human herpesvirus 6 variants A a

A:Reference number: Z22732; MUID:99412319; PMID:10482554

A:Accession: T43979

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-389 <ISE>

A:Cross-references: EMBL:AB021506; NID:g4995977; PIDN:BA78240.1; PID:g4996007

A:Experimental source: strain HST; pop. variant B

R:Domínguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.

J. Virol. 73, 8040-8052, 1999

A:Title: Human herpesvirus 6B genome sequence: coding content and comparison with hu

A:Reference number: Z22734; MUID:99412318; PMID:10482553

A:Accession: T44165

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-389 <DOM>

A:Cross-references: EMBL:AF157706; PIDN:AAD49631.1

A:Experimental source: strain Z29; variant B

C:Genetics:

A:Gene: U19

C:Superfamily: human herpesvirus 6 hypothetical protein U19

Query Match 42.9%; Score 42; DB 2; Length 389;

Best Local Similarity 33.3%; Pred. No. 52;

Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 4 VALYGLKQDNGVSA 18

||||: ||: |

Db 79 MSFMGVTEWEGASA 93

RESULT 29

T09313

Immediate-early protein 4 - human herpesvirus 6 (strain U1102)

C:Species: human herpesvirus 6

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 28-Jul-2000

C:Accession: T09313

R:Nicholas, J.; Martin, M.

J. Virol. 68, 597-610, 1994

A:Title: Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of

A:Reference number: Z16644; MUID:94118404; PMID:8289364

A:Accession: T09313

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-412 <NIC>

A:Cross-references: EMBL:L25528; NID:g451932; PIDN:AAA16726.1; PID:g451944

C:Genetics:

A:Gene: E1LF4

C:Superfamily: human herpesvirus 6 hypothetical protein U19

Query Match 42.9%; Score 42; DB 2; Length 412;

Best Local Similarity 33.3%; Pred. No. 55;

Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 4 VALYGLKQDNGVSA 18

||||: ||: |

Db 102 MSFMGVTEWEGASA 116

RESULT 30

AH2117

hypothetical protein alr2495 [Imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AH2117

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AH2117

A:Status: preliminary

A:Gene: xynE
C:Function:
A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans
A:Pathway: xylan degradation
C:Superfamily: Pseudomonas endo-1,4-beta-xylanase E; endo-1,4-beta-xylanase homology;
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-661/Product: endo-1,4-beta-xylanase E #status predicted <MAT>
F:740-226/Domain: endo-1,4-beta-xylanase homology <XYL>
F:403-549/Domain: nodB homology <NOdB>
F:616-653/Domain: glycosidase GWGW domain homology <GWG>
F:116,213/Active site: Glu #status predicted

Query Match 42.9%; Score 42; DB 1; Length 661;
Best Local Similarity 87.5%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 QDNWGVSA 18
|||||
Db 522 QDNWGASA 529

RESULT 33
B75622
Hypothetical protein - Deinococcus radiodurans (strain RI)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: B75622
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, F.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S. Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: B75622
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-691 <WHI>
A:Cross-references: GB:AE001826; NID:g6460827; PIDN:g6460924; TIGR:DR
A:Experimental source: Strain RI
C:Genetics:
A:Gene: DRB0037
A:Map position: megaplasmid
A:Genome: Plasmid
A>Note: plasmid MP1

Query Match 42.9%; Score 42; DB 2; Length 691;
Best Local Similarity 50.0%; Pred. No. 93;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNWGVSA 18
||| | | | : |
Db 310 NPATAATYELVPLDLSGLPA 327

RESULT 34
A47296
Thiazide-sensitive electroneutral sodium-chloride cotransporter - winter flounder
C:Species: Pseudeurostectes americanus (winter flounder)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 02-Mar-2001
C:Accession: A47296
R:Gamoa, G.; Saltzberg, S.N.; Lombardi, M.; Miyanoshta, A.; Lytton, J.; Hediger, M.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 2749-2753, 1993
A:title: Primary structure and functional expression of a cDNA encoding the thiazide-
A:Reference number: A47296; MUID:93219361; PMID:8464884
A:Accession: A47296
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1023 <GAM>
A:Experimental source: urinary bladder
C:Note: sequence extracted from NCBI backbone (NCBI:128719, NCBIP:128720)
C:Superfamily: rat bumetanide-sensitive Na+/K+Cl-cotransport protein

Query Match 42.9%; Score 42; DB 2; Length 1023;
Best Local Similarity 58.3%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 PTVALYGLKQDW 13
| | | | | | | |
Db 727 PNVLGMGPKDW 738

RESULT 35

C64474
hypothetical protein MJ1396 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: C64474
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: C64474
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2894 <BUL>
A:Cross-references: GB:U67579; GB:L77117; NID:gl592037; PIDN:AAB99406.1; PID:gl592043; T
C:Genetics:
A:Map position: FOR1347709-1356393

Query Match 42.9%; Score 42; DB 2; Length 2894;
Best Local Similarity 57.1%; Pred. No. 4.1e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWN 14
| | | | | | | |
Db 884 NSTIALYGLVDAYN 897

Search completed: October 6, 2003, 07:49:35
Job time : 14.6716 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:02 ; Search time 5.67164 Seconds
(without alignments)
165.831 Million cell updates/sec

Title: US-09-765-739A-1

Perfect score: 110

Sequence: 1 KSTGVGFLKHDWDGSPILK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	50	45.5	756	1 K6PF_CABEL	Q27483 caenorhabdi
2	46	41.8	1182	1 CGA2_HELPY	P55746 helicobacte
3	45	40.9	570	1 ILVD_LACLA	Q02139 lactococcus
4	45	40.9	704	1 DPOL_BPT3	P20311 bacterioph
5	45	40.9	704	1 DPOL_BPT7	P00581 bacterioph
6	44	40.0	160	1 MENG_DEIRA	Q9rx10 deinococcus
7	44	40.0	368	1 Y093_HAEIN	P44509 haemophilus
8	43	39.1	255	1 YABD_BACSU	P37545 bacillus su
9	43	39.1	1186	1 CAGA_HELPY	P55980 helicobacte
10	42.5	38.6	84	1 NI9M_HUMAN	O95167 homo sapien
11	42	38.2	347	1 NU2M_BOVIN	P03892 bos taurus
12	42	38.2	347	1 NU2M_HIPAM	Q92220 hippopotamu
13	42	38.2	347	1 NU2M_SHEEP	O78748 ovis aries
14	42	38.2	475	1 MTHC_DRONE	P83119 drosophila
15	42	38.2	3344	1 POLG_PRSVH	Q01901 p genome po
16	41	37.3	102	1 YMSA_RHIME	Q07602 rhizobium m
17	41	37.3	216	1 TRPE_METKA	O8txz9 methanopyru
18	41	37.3	414	1 KC12_RAT	O62762 rattus norv
19	41	37.3	415	1 KC12_HUMAN	P78368 homo sapien
20	41	37.3	1081	1 AT18_HUMAN	Q8t660 homo sapien
21	40	36.4	203	1 LIPE_BUCBP	O89a18 buchnera ap
22	40	36.4	221	1 TRPE_HALVO	P52563 halobacteri
23	40	36.4	246	1 CYSQ_ECOLI	P22555 escherichia
24	40	36.4	326	1 GBLP_NICPL	P93340 nicotiana p
25	40	36.4	326	1 GBLP_TOBAC	P49026 nicotiana t
26	40	36.4	377	1 YA67_METTH	O27139 methanobact
27	40	36.4	483	1 PEPI_HAEIN	P44817 haemophilus
28	40	36.4	567	1 UREL_PROMI	P17086 proteus mir
29	40	36.4	578	1 NADB_ANASP	O8yx16 anabaena sp
30	40	36.4	600	1 YB14_MYCPN	P75448 mycoplasma
31	40	36.4	604	1 NET1_HUMAN	O95631 homo sapien
32	40	36.4	604	1 NET1_MOUSE	O09118 mus musculu
33	40	36.4	606	1 NET1_CHICK	Q09922 gallus gall

Q9pf41 xylella fas	792	1	OSTA_XYLFA
Q8pes0 xanthomonas	885	1	PLSB_XANAC
Q9plb0 chlamydia m	1520	1	PMPD_CHLMU
T2925 petunia hyb	419	1	CHSD_PETHY
Q8y9p2 bruceella me	518	1	NAEK_BRUME
Q8fwn5 bruceella su	525	1	NAEK_BRUSE
Q8vpc0 drosophila	952	1	KP58_DROME
Q8y1k8 anabaena sp	92	1	PAK2_ANASP
Q48650 salix bakko	113	1	SUIL_SALBA
Q8prx4 methanosarc	226	1	TRE2_METMA
P45908 bacillus su	284	1	YQAK_BACSU
O30207 archaeoglob	371	1	Y028_ARCFU
P43416 streptomyce	437	1	SECY_STRSC
P39045 actinomadr	538	1	DAC_ACTSP
Q9aj06 arthrobacte	575	1	TREZ_ARTRM
P36150 saccharomyc	593	1	SUMT_YEAST
P11129 bacterioph	648	1	VP3_BPPH6
Q8pce0 xanthomonas	809	1	OSTA_XANCP
O8pp22 xanthomonas	813	1	OSTA_XANAC
P33775 saccharomyc	817	1	PMT1_YEAST
O61309 lymnaea sta	1153	1	NOS_LYMST
P34203 african swi	1191	1	TOP2_ASFM2
Q00942 african swi	1192	1	TOP2_ASFB7
O28768 archaeoglob	201	1	YF04_ARCFU
P26190 porcine rot	1082	1	RRPO_ROTFC
Q941v4 arabidopsis	113	1	SUI2_ARATH
Q99a85 mus musculu	196	1	R18A_MOUSE
Q921k8 helicobacte	282	1	BI0B_HELPJ
O25956 helicobacte	289	1	BI0B_HELPY
O8yvc1 anabaena sp	345	1	AROE_ANASP
P13157 turkey herp	310	1	KITH_HSVTF
O42248 brachydanio	317	1	GBLP_BRARE
P25388 homo sapien	320	1	GBLP_HUMAN
Q9h255 homo sapien	320	1	OXE2_HUMAN
O86288 rattus norv	320	1	OXE2_RAT
P08186 escherichia	322	1	PTNA_ECOLI
P25987 turkey herp	350	1	KITH_HSVTU
Q9d9q7 tetraodon f	377	1	CC37_TETFL
O61081 mus musculu	379	1	CC37_MOUSE
P63692 rattus norv	379	1	CC37_RAT
P00973 homo sapien	400	1	OAS1_HUMAN
P45817 yarrowia li	404	1	PEX9_YARLI
P32140 escherichia	413	1	YIHS_ECOLI
O54435 rahnella aq	415	1	SACB_RAHAA
Q58008 methanococc	430	1	Y588_METJA
P54968 arabidopsis	442	1	ILR1_ARATH
Q9uy49 pyrococcus	485	1	IMDL_PYRAB
P05429 synecocyst	507	1	PSBB_SYNY3
P18314 klebsiella	567	1	UREI_KLEAE
P36729 human papil	609	1	VE1_HPV49
O74343 schizosacch	680	1	YH2X_SCHPO
Q60411 cavia porce	735	1	AD02_CAVPO
Q9ydx6 aeopyrum p	815	1	AOX1_AERPE
P28962 equine herp	1081	1	ULS2_HSVB
O57861 pyrococcus	1431	1	DP2L_PYRHO
P37800 toscana vir	2095	1	RRPL_TOSV
P01549 streptomyce	144	1	MACM_STRMA
O69600 mycobacteri	345	1	ALF_MYCLE
Q9pl12 chlamydia m	352	1	TRMD_CHLMU
P03441 influenza a	550	1	HEMA_IABAN
P12589 influenza a	550	1	HEMA_IAME6
P03440 influenza a	566	1	HEMA_IAEN7
P26139 influenza a	566	1	HEMA_IAZCO
P26141 influenza a	566	1	HEMA_IAZUK
P03435 influenza a	567	1	HEMA_IANV1
O94039 candida alb	677	1	TKT1_CANAL
Q02510 staphylococ	688	1	LIP_STAEP

ALIGNMENTS

RESULT 1

```
K6PF_CAEEL
ID K6PF_CAEEL STANDARD; PRT; 756 AA.
AC Q27483;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)
DE (Phosphohexokinase).
GN C50P4.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -!- PATHWAY: Key control step of glycolysis.
CC -!- SIMILARITY: Belongs to the phosphofructokinase family. Two domains
CC subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z70750; CAA94737.1; -.
DR PIR; T20109; T20109.
DR HSSP; P00512; 3PFK.
DR WormPep; C50P4.2; CE05467.
DR InterPro; IPR000023; Pfpruckkinase.
DR Pfam; PF00365; PFK; 2.
DR PRINTS; PR00476; PFRCTKINASE.
DR PRODOM; PD000707; Pfructkinase; 2.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
DR Kinase; Transferase; Glycolysis; Repeat.
KW Kinase.
SQ SEQUENCE 756 AA; 83301 MW; 26A9B801D286534 CRC64;
Query Match 45.5%; Score 50; DB 1; Length 756;
Best Local Similarity 57.1%; Pred. No. 2.7;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Qy 2 STVGVLGKHKHDWDG 15
I : | | | | |
Db 432 SGIOVIGIKHGDG 445
RESULT 2
CGA2_HELPY
ID CGA2_HELPY STANDARD; PRT; 1182 AA.
AC P55746;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
DE CAGA OR CAI.
GN Helicobacter pylori (Campylobacter pylori).
OS Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 53726 / 84-183;
RA Tummuru M.K.R.; Cover T.L.; Blaser M.J.;
RA "Cloning and expression of a high-molecular-mass major antigen of
RT Helicobacter pylori: evidence of linkage to cytotoxin production.";
```

```
RL Infect. Immun. 61:1799-1809(1993).
CC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11714; -. NOT_ANNOTATED_CDS.
DR InterPro; IPR005169; Caga.
DR InterPro; IPR004355; IVSec_caga.
DR Pfam; PF03507; Caga; 1.
DR PRINTS; PR01553; TYPE4SSCAGA.
DR Antigen.
KW DOMAIN.
SQ SEQUENCE 878 AA; 885 POLY-ASN.
CC SEQUENCE 1182 AA; 131503 MW; C916817E2EE57BB4 CRC64;
Query Match 41.8%; Score 46; DB 1; Length 1182;
Best Local Similarity 58.3%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 5 GVFGKLDWDGS 16
I | | | | |
Db 323 GGFGAKHDWNAT 334
RESULT 3
ILVD_LACLA
ID ILVD_LACLA STANDARD; PRT; 570 AA.
AC Q02139;
DT 01-JUL-1993 (Rel. 26, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD).
GN ILVD OR LLI223.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCDO 2118;
RX MEDLINE=93015710; PubMed=1400210;
RA Godon J.-J., Chopin M.-C., Ehrlich S.D.;
RT "Branched-chain amino acid biosynthesis genes in Lactococcus lactis
RT subsp. lactis.";
RL J. Bacteriol. 174:6580-6589(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -!- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-
CC oxobutanoate + H(2)O.
CC -!- COFACTOR: BINDS 1 4Fe-4S CLUSTER (POTENTIAL).
CC -!- PATHWAY: Valine and isoleucine biosynthesis; fourth step.
CC -!- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; U92974; AAB81918.1; -.
DR EMBL; AE006354; AAK05321.1; -.
DR PIR; G86777; G86777.
DR PIR; S35137; S35137.
DR HAMAP; MF_00012; -. 1.
DR InterPro; IPR004404; ILVD.
DR InterPro; IPR000581; ILVD_EDD_family.
DR Pfam; PF00920; ILVD_EDD; 1.
DR ProDom; PD002691; ILVD_EDD_family; 1.
DR TIGRFAMs; TIGR00110; ILVD; 1.
DR PROSITE; PS00886; ILVD_EDD_1; 1.
DR PROSITE; PS00887; ILVD_EDD_2; 1.
DR Branched-chain amino acid biosynthesis; Lyase; Iron; Iron-sulfur;
KW 4Fe-4S; Complete proteome.
FT METAL 135 135 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 207 207 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT CONFLICT 23 25 QDP -> PRS (IN REF. 1).
FT CONFLICT 179 179 S -> A (IN REF. 1).
FT CONFLICT 246 246 E -> D (IN REF. 1).
FT CONFLICT 313 313 T -> I (IN REF. 1).
FT CONFLICT 530 530 D -> N (IN REF. 1).
SQ SEQUENCE 570 AA; 60737 MW; 91D45384FD9445D3 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 570;
Best Local Similarity 47.1%; Pred. No. 13;
Matches 8; Conservative 4; Mismatches 3; Indels 2; Gaps 1;

QY 1 KSTVGVLGLKHDWDSP 17
|: ||: : |||||
DB 46 KAQVGIVSM--DWDGNP 60

RESULT 4
DPOL_BPT3 STANDARD; PRT; 704 AA.
AC P20311;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN 5.
OS Bacteriophage T3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Luria;
RX MEDLINE=90133923; PubMed=2614843;
RA Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RT "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.";
RL J. Mol. Biol. 210:687-701(1989).
CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS A 3' TO 5' EXONUCLEASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -!- SUBUNIT: COMPOSED OF TWO SUBUNITS. ONE IS ENCODED BY THE PHAGE AND
CC THE OTHER IS ENCODED BY THE HOST THIOREDOXIN.
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17255; CAA35140.1; -.
DR PIR; S07512; S07512.
DR HSSP; P00581; 1T7P.
DR InterPro; IPR001098; DNA_pol.
DR InterPro; IPR002298; DNA_poli.
```

```
DR Pfam; PF00476; DNA_pol_A; 1.
DR PRINTS; PR00868; DNAPOLI.
DR SMART; SM00482; POLAC; 1.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Hydrolase; Exonuclease.
SQ SEQUENCE 704 AA; 79985 MW; D65DFBD99AE31234 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 704;
Best Local Similarity 87.5%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 GLKHDWDG 15
|||||
DB 637 GLKHGWDG 644

RESULT 5
DPOL_BPT7 STANDARD; PRT; 704 AA.
AC P00581;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) (T7 DNA polymerase).
GN 5.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10760;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RA Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements.";
RL J. Mol. Biol. 166:477-535(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84164887; PubMed=6708104;
RA Moffatt B.A., Dunn J.J., Studier F.W.;
RT "Nucleotide sequence of the gene for bacteriophage T7 RNA
RT polymerase.";
RL J. Mol. Biol. 173:265-269(1984).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE=98101638; PubMed=9440688;
RA Doublie S., Tabor S., Long A.M., Richardson C.C., Ellenberger T.;
RT "Crystal structure of a bacteriophage T7 DNA replication complex at
RT 2.2-A resolution.";
RL Nature 391:251-258(1998).
CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS A 3' TO 5' EXONUCLEASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -!- SUBUNIT: COMPOSED OF TWO SUBUNITS. ONE IS ENCODED BY THE PHAGE AND
CC THE OTHER IS ENCODED BY THE HOST THIOREDOXIN.
CC -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; V01146; CAA24412.1; -.
DR PIR; A00716; DJBPT7.
DR PDB; 1T7P; 25-FEB-98.
DR InterPro; IPR001098; DNA_pol.
DR InterPro; IPR002298; DNA_poli.
DR Pfam; PF00476; DNA_pol_A; 1.
```

```
DR PRINTS: PR00868; DNAPOLI.
DR SMART: SM00482; POLAC; I.
DR PROSITE: PS00447; DNA_POLYMERASE_A; 1.
DR Transferrase; DNA-directed DNA polymerase; DNA replication; Hydrolase;
KW Exonuclease; DNA-binding; 3D-structure.
SQ SEQUENCE 704 AA; 79691 MW; 17089CE2AD9FB596 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 704;
Best Local Similarity 87.5%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 GLKHDWDG 15
      |||||
Db 637 GLKHGWDG 644

RESULT 6
ID MENG_DEIRA STANDARD; PRT; 160 AA.
AC Q9PWL0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable S-adenosylmethionine:2-demethylmenaquinone methyltransferase
DE (EC 2.1.1.-.-).
GN MENG OR DR0859.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RL SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: CONVERTS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK)
CC (by similarity)
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE MENG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE001940; AAF10437.1; -.
CC DR PIR: A75466; A75466.
CC DR TIGR: DR0859; -.
CC DR HAMAP: MF_00471; -.
CC DR InterPro: IPR005493; Methyltransf_6.
CC DR Pfam: PF03737; Methyltransf_6; 1.
CC DR Menaquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SEQUENCE 160 AA; 16860 MW; 54D6F226CA19EC0E CRC64;

Query Match 40.0%; Score 44; DB 1; Length 160;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 4 VGVFGLKHDWDG 15
      :||||:|
Db 4 VGVFGLKHDWDG 15

us-09-765-739a-1.rsp
79 LGVFCVNGWEG 90

Db 79 LGVFCVNGWEG 90

RESULT 7
ID Y093_HAEIN STANDARD; PRT; 368 AA.
AC P44509;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0093.
GN HI0093.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RL SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Klenk A.P., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Spriggs T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE CDAR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32695; AAC21771.1; -.
CC DR PIR: E64142; E64142.
CC DR TIGR: HI0093; -.
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 368 AA; 42251 MW; C77F1C9EF043B89A CRC64;

Query Match 40.0%; Score 44; DB 1; Length 368;
Best Local Similarity 46.2%; Pred. No. 12;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 8 GLKHDWDGSPILK 20
      |||||:|
Db 283 GLSHSQGNELIK 295

RESULT 8
ID YABD_BACSU STANDARD; PRT; 255 AA.
AC P37545;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative deoxyribonuclease yabd (EC 3.1.21.-).
GN YABD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RL SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA
```

ID	CAGA_HELPY	STANDARD;	PRT;	1186 AA.
AC	P55980;			
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Cytotoxicity associated immunodominant antigen (120 kDa protein)			
GN	(CAGA pathogenicity island protein 26).			
GN	CAGA OR CAI OR CAG26 OR HP0547.			
OS	Helicobacter pylori (Campylobacter pylori).			
OS	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;			
OC	Helicobacteraceae; Helicobacter.			
NCBI_TaxID=210;				
[1]				
RN	SEQUENCE FROM N.A.			
RP	STRAIN=26695 / ATCC 700392;			
RC	MEDLINE=97394467; PubMed=9252185;			
RX	Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,			
RA	Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,			
RA	Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,			
RA	Lofthus B., Richardson D., Dodson R., Khalak M.G., Glodek A.,			
RA	McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,			
RA	Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,			
RA	Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,			
RA	Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,			
RA	Venter J.C.;			
RT	"The complete genome sequence of the gastric pathogen Helicobacter			
RT	pylori.";			
RL	Nature 388:539-547(1997).			
CC	- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,			
CC	OR FUNCTION OF THE CYTOTOXIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AE000569; AAD07614.1; -			
DR	PIR; C64588; C64588.			
DR	TIGR; HP0547; -			
DR	InterPro; IPR005169; Caga.			
DR	InterPro; IPR004355; IVSec_caga.			
DR	Pfam; PF03507; Caga; 1.			
DR	PRINTS; PR01553; TYPE4SSCAGA.			
KW	Antigen; Complete proteome.			
FT	DOMAIN 247 250 POLY-THR.			
FT	DOMAIN 883 889 POLY-ASN.			
SQ	SEQUENCE 1186 AA; 132386 MW; B05C3F2CCC444F4 CRC64;			
Query Match	39.1%; Score 43; DB 1; Length 1186;			
Best Local Similarity	58.3%; Pred. No. 60;			
Matches	7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;			
Qy	5 GFVEGLKHDNDGS 16			
Dd	I I I I I I I ;			
	328 GGFGDKHWNAT 339			
RESULT 10				
NI9M_HUMAN				
ID	NI9M_HUMAN	STANDARD;	PRT;	84 AA.
AC	O95167;			
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	NADH+ubiquinone oxidoreductase B9 subunit (EC 1.6.5.3) (EC 1.6.99.3)			
DE	(Complex I-B9) (CI-B9).			
GN	NDUFA3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99097250; PubMed=9878551;
RA Loeffen J.L.C.M., Triepels R.H., van den Heuvel L., Schuelke M.,
RA Buskens C.A.F., Smeets R.J.P., Trijbels J.M.F., Smeitink J.A.M.;
RT "cDNA of eight nuclear encoded subunits of NADH-ubiquinone
RT oxidoreductase; human complex I cDNA characterization completed.";
RL Biochem. Biophys. Res. Commun. 253:415-422(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Blood;
RA MEDLINE=20493367; PubMed=11042152;
RA Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,
RA Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,
RA Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;
RT "Cloning and functional analysis of cDNAs with open reading frames for
RT 300 previously undefined genes expressed in CD34+ hematopoietic
RT stem/progenitor cells.";
RL Genome Res. 10:1546-1560(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Prostate;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Basha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Prange C.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gnaratine P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gagliardi J.,
RA Villalón D.K., Muzley D.M., Sodergren E.J., Li X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Butterfield Y.S., Grimwood J., Schmutz J., Myers R.M.,
RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
CC TO BE UBIQUINONE.
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- CATALYTIC ACTIVITY: NADH + acceptor = NAD(+) + reduced acceptor.
CC -!- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I B9 SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: AF044955; AAD05420.1; -
CC EMBL: AF070653; AAD20959.1; -
CC EMBL: BC022369; AAB22369.1; -
CC PIR: JE0379; JF0379.
CC GenBank: U00003.7686; NDUPA3.
CC MIM: 603832.
CC GO: GO:0005748; C:NADH dehydrogenase complex (ubiquinone) (se... TAS.
CC GO: GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; TAS.
CC Oxidoreductase; Ubiquinone; NAD; Mitochondrion; Transmembrane.
CC TRANSMEM 19 39 POTENTIAL.
CC SEQUENCE 84 AA; 9279 MW; 38B27A96D7A05D31 CRC64;
DR EMBL: AF044955; AAD05420.1; -
DR EMBL: AF070653; AAD20959.1; -
DR EMBL: BC022369; AAB22369.1; -
DR PIR: JE0379; JF0379.
DR GenBank: U00003.7686; NDUPA3.
DR MIM: 603832.
DR GO: GO:0005748; C:NADH dehydrogenase complex (ubiquinone) (se... TAS.
DR GO: GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; TAS.
KW Oxidoreductase; Ubiquinone; NAD; Mitochondrion; Transmembrane.
FT TRANSMEM 19 39 POTENTIAL.
SQ SEQUENCE 84 AA; 9279 MW; 38B27A96D7A05D31 CRC64;
Query Match 38.6%; Score 42.5; DB 1; Length 84;
Best Local Similarity 56.2%; Pred. No. 4.6;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;
QY 4 VCVFGLKHDWGSPIL 19
DB 5 VGAF-LKNWDXEPVL 19
RESULT 11
NUZM_BOVIN
ID NUZM_BOVIN STANDARD; PRT; 347 AA.
AC P03892; Q8SF93;
DC 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN MTND2 OR ND2.
OS Bos taurus (Bovine).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Heart;
RA MEDLINE=83010260; PubMed=7120390;
RA Anderson S., de Bruijn M.H.L., Coulson A.R., Eperon I.C., Sanger F.,
RA Young I.G.;
RT "Complete sequence of bovine mitochondrial DNA. Conserved features of
RT the mammalian mitochondrial genome.";
RL J. Mol. Biol. 156:683-717(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=65, 66, D, and F;
RA Wetstein P.J.;
RT "Bos taurus mitochondrial protein coding regions.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: V00654; CAA23998.1; -
CC EMBL: J01394; AAB59269.1; -
CC EMBL: AF490528; AAM08323.1; -
CC EMBL: AF490529; AAM08336.1; -
CC EMBL: AF493541; AAM12790.1; -
CC EMBL: AF493542; AAM12803.1; -
CC PIR: A00415; OXBO2M.
CC InterPro: IPR001750; Oxidored_q1.
CC Pfam: PF0361; oxidored_q1.1.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
FT VARIANT 227 227 T -> I (IN STRAIN F).
SQ SEQUENCE 347 AA; 39254 MW; 24105538A1374585 CRC64;
Query Match 38.2%; Score 42; DB 1; Length 347;
Best Local Similarity 38.9%; Pred. No. 24;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
QY 2 STGVFGLKHDWGSPIL 19
DB 223 STTTLSLSHTWTKPTIM 240

```

```

RESULT 12
NU2M_HIPAM
AC Q92220; STANDARD; PRT; 347 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN MTND2 OR ND2 OR NADH2.
OS Hippopotamus amphibius (Hippopotamus).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Hippopotamidae; Hippopotamus.
OX NCBI_TaxID=9833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99098146; PubMed=9881471;
RA Ursing B.M., Arnason U.;
RT "Analyses of mitochondrial genomes strongly support a hippopotamus-
RT whale clade.";
RL Proc. R. Soc. Lond., B, Biol. Sci. 265:2251-2255(1998).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF010406; RAD10097.1; -
DR F01051; T11051.
DR InterPro: IPR003917; NADH_oxred2.
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam: PF00361; oxidored_q1; 1.
DR PRINTS: PR01436; NADHDHGNASE2.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
SQ SEQUENCE 347 AA; 39128 MW; 520DF0A0D6C991B CRC64;

Query Match 38.2%; Score 42; DB 1; Length 347;
Best Local Similarity 38.9%; Pred. No. 24;
Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSPIL 19
|| | | | |
DB 223 STTTLSLSTWTKAPII 240

RESULT 14
MTHC_DROME
ID MTHC_DROME STANDARD; PRT; 475 AA.
AC P83119;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Probable G-protein-coupled receptor Mth-like 12 precursor (Methuselah-
DE like 12 protein).
GN MTHL12 OR MTH-LIKE-12.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merklov G., Milshina N.V., Mobarry J., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zeng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP IDENTIFICATION.
 RX MEDLINE=21173629; PubMed=11274391;
 RA West A.P. Jr., Llamas L.L., Snow P.M., Benzer S., Bjorkman P.J.;
 RA "Crystal structure of the ectodomain of Methuselah, a Drosophila G
 RT protein-coupled receptor associated with extended lifespan.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3744-3749(2001).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED
 CC RECEPTORS. MTH SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE003699; ; NOT_ANNOTATED_CDS.
 DR FlyBase; FBgn0045442; mth12.
 DR GO; GO:0004930; P:G-protein coupled receptor activity; ISS.
 DR GO; GO:0008340; P:determination of adult life span; ISS.
 DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . ; ISS.
 DR GO; GO:0006950; P:response to stress; ISS.
 DR InterPro; IPR000832; GPCR_secretin.
 DR Pfam; PF00002; 7tm2; 1.
 DR PROSITE; PS00649; G-PROTEIN_RECEP_F2_1; FALSE_NEG.
 DR PROSITE; PS00650; G-PROTEIN_RECEP_F2_2; FALSE_NEG.
 DR PROSITE; PS0261; G-PROTEIN_RECEP_F2_4; 1.
 KW Receptor; G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Signal; Multigene family.
 FT SIGNAL 1 17
 FT CHAIN 18 475
 FT PROBABLE G-PROTEIN-COUPLED RECEPTOR MTH-
 FT LIKE 12.
 FT DOMAIN 18 214
 FT TRANSMEM 215 235
 FT DOMAIN 236 242
 FT TRANSMEM 243 263
 FT DOMAIN 264 275
 FT TRANSMEM 276 296
 FT DOMAIN 297 307
 FT TRANSMEM 308 328
 FT DOMAIN 329 360
 FT TRANSMEM 361 381
 FT DOMAIN 382 403
 FT TRANSMEM 404 424
 FT DOMAIN 425 442
 FT TRANSMEM 443 463
 FT DOMAIN 464 475
 FT CYTOPLASMIC (POTENTIAL).
 FT BY SIMILARITY.
 FT DISULFID 27 81
 FT BY SIMILARITY.
 FT DISULFID 83 88
 FT BY SIMILARITY.
 FT DISULFID 92 183
 FT BY SIMILARITY.
 FT DISULFID 93 104
 FT BY SIMILARITY.
 FT DISULFID 149 203
 FT BY SIMILARITY.
 FT CARBOHYD 19 19
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 34 34
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 475 AA; 55424 MW; 3530FF875F77264 CRC64;
 Query Match 38.2%; Score 42; DB 1; Length 475;
 Best Local Similarity 61.5%; Pred. No. 34;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 4 VGVFGLKHDWGS 16
 II III II:
 DB 342 VGYFGLTDWNS 354
 II III II:
 RESULT 15
 POLG PRSVH
 ID POLG PRSVH STANDARD; PRT; 3344 AA.
 AC Q01901;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome Polyprotein [Contains: N-terminal protein (p1); Helper
 DE Component Proteinase (EC 3.4.22.45) (HC-Pro); Protein P3; 6 kDa
 DE Proteinase (6K1); Cytoplasmic inclusion protein (CI); 6 kDa protein 2
 DE (6K2); Genome-linked protein (VPG); Nuclear inclusion protein A (NI-A)
 DE (NIA) (EC 3.4.22.44) (49 kDa proteinase) (49 kDa-Pro); Nuclear
 DE inclusion protein B (NI-B) (NIB) (RNA-directed RNA polymerase)
 DE (EC 2.7.7.48); Coat protein (CP)].
 OS Papaya ringspot virus (strain P / mutant HA).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
 OC Potyvirus.
 CC Potyvirus.
 CC NCBI_TaxID=31731;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wang C.H., Bau H.J., Yeh S.D.;
 RT "Comparison of the nuclear inclusion b protein and coat protein genes
 RT of five papaya ringspot virus strains distinct in geographic origin
 RL and pathogenicity.";
 RN Phytopathology 84:1205-1210(1994).
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93019006; PubMed=1402799;
 RA Yeh S.D., Jan F.J., Chiang C.H., Doong T.J., Chen M.C.;
 RA Chung P.H., Bau H.J.;
 RT "Complete nucleotide sequence and genetic organization of papaya
 RT ringspot virus N.A.";
 RL J. Gen. Virol. 73:2531-2541(1992).
 RP SEQUENCE OF 2561-3344 FROM N.A.
 RX MEDLINE=93090098; PubMed=1456896;
 RA Wang C.H., Yeh S.D.;
 RT "Nucleotide sequence comparison of the 3'-terminal regions of severe-,
 mild, and non-papaya infecting strains of papaya ringspot virus.";
 RL Arch. Virol. 127:345-354(1992).
 CC -1- FUNCTION: HELPER COMPONENT. PROTEINASE IS REQUIRED FOR APHID
 CC -1- TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
 CC -1- FUNCTION: CYTOPLASMIC INCLUSION. PROTEIN HAS HELICASE ACTIVITY. IT
 CC MAY BE INVOLVED IN REPLICATION.
 CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: Hydrolyzes glutamyl bonds, and activity is
 CC further restricted by preferences for the amino acids in P6 - P1'
 CC that vary with the species of potyvirus, e.g. Glu-Xaa-Xaa-Tyr-Xaa-
 CC Gln+(Ser or Gly) for the enzyme from tobacco etch virus. The
 CC natural substrate is the viral polyprotein, but other proteins and
 CC oligopeptides containing the appropriate consensus sequence are
 CC also cleaved.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(N).
 CC -1- CATALYTIC ACTIVITY: Hydrolyzes a Gly-I-Gly bond at its own C-
 CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-I-Gly, in the
 CC processing of the potyviral polyprotein.
 CC -1- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
 CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE

CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
 CC INDIVIDUAL PROTEINS.
 CC -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
 CC -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
 CC -!- SIMILARITY: BELONGS TO THE POTVIRUSES POLYPROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X67673; CAA47905.1; -;
 DR EMBL: S46722; AAB23789.1; -;
 DR EMBL: X67672; CAA47904.1; -;
 DR PIR: JQ1899; JQ1899.
 DR MEROPS: C04.009; -;
 DR MEROPS: C06.001; -;
 DR MEROPS: S30.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR001730; Peptidase_C4.
 DR InterPro: IPR001456; Peptidase_C6.
 DR InterPro: IPR001592; Poty_coat.
 DR InterPro: IPR002540; Poty_P1.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00863; Peptidase_C4; 1.
 DR Pfam: PF00851; Peptidase_C6; 1.
 DR Pfam: PF00767; Poty_coat; 1.
 DR Pfam: PF01577; Poty_P1; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR PRINTS: PR00966; NIAPOTYPTASE.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 KW Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
 KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
 KW ATP-binding.

FT CHAIN 1 529 N-TERMINAL PROTEIN (BY SIMILARITY).
 FT CHAIN 530 1149 HELPER COMPONENT PROTEINASE
 FT (BY SIMILARITY).
 FT CHAIN 1150 ? PROTEIN P3 (BY SIMILARITY).
 FT CHAIN ? 1401 6 kDa PROTEIN 1 (BY SIMILARITY).
 FT CHAIN 1402 2036 CYTOPLASMIC INCLUSION PROTEIN (BY
 FT SIMILARITY).
 FT CHAIN 2037 2093 6 kDa PROTEIN 2 (BY SIMILARITY).
 FT CHAIN 2094 ? GENOME-LINKED PROTEIN (BY SIMILARITY).
 FT CHAIN ? 2520 NUCLEAR INCLUSION PROTEIN A
 FT (BY SIMILARITY).
 FT CHAIN 2521 3037 NUCLEAR INCLUSION PROTEIN B
 FT (BY SIMILARITY).
 FT CHAIN 3038 3344 COAT PROTEIN (BY SIMILARITY).
 FT BINDING 2156 2156 COVALENT LINKAGE OF VIRAL RNA (BY
 FT SIMILARITY).
 FT NP_BIND 1486 1493 ATP (POTENTIAL).
 FT SEQUENCE 3344 AA; 381040 MW; E90CD7523AC5243D CRC64;

Query Match 38.2%; Score 42; DB 1; Length 3344;
 Best Local Similarity 61.5%; Pred. No. 2.6e+02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 GLKWDGSPILK 20
 ||| |||| :|
 Db 2049 GLKWDGSLMTK 2061

RESULT 16
 YMSA_RHIME

ID YMSA_RHIME STANDARD; PRT; 102 AA.
 AC Q07602;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 11.0 kDa protein in mosa 5' region (ORF1).
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=L5-30;
 RX MEDLINE=93352426; PubMed=8349559;
 RA Murphy P.J., Trenz S.P., Grzemski W., de Bruijn F.J., Schell J.;
 RT "The Rhizobium meliloti rhizopine mos locus is a mosaic structure
 RT facilitating its symbiotic regulation.";
 RL J. Bacteriol. 175:5193-5204(1993).
 CC -!- FUNCTION: THIS PROTEIN IS EITHER NOT EXPRESSED, EXPRESSED AT LOW
 CC LEVELS OR RAPIDLY DEGRADED.
 CC -!- SIMILARITY: TO THE N-TERMINAL OF NITROGENASE IRON PROTEIN (NIFH).
 CC -!- HAS LOST THE ATP-BINDING SITE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L17071; AAA26300.1; -;
 DR PIR: A53308; A53308.
 KW Hypothetical protein.
 SQ SEQUENCE 102 AA; 11047 MW; 0849F4FC0BA022A8 CRC64;
 Query Match 37.3%; Score 41; DB 1; Length 102;
 Best Local Similarity 50.0%; Pred. No. 9.8;
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 9 LKHDWDGSPILK 20
 :||| | | | |
 Db 71 MKHKWKGQPLPK 82
 RESULT 17
 TRPF_METKA
 ID TRPF_METKA STANDARD; PRT; 216 AA.
 AC Q8TXZ9;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI).
 GN TRPF OR MK0508.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
 OC Methanopyrus
 OX NCBI_TaxID=2320;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=AV19 / DSM 6324 / JCM 9639;
 RX MEDLINE=21927647; PubMed=11930014;
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
 RA Nale D.A., Rogozin I.B., Tatusov R.D., Wolf Y.I., Stetter K.O.,
 RA Malykh A.G., Koonin E.V., Kozavkin S.A.;
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
 RT and monophyly of archaeal methanogens.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 CC -!- CATALYTIC ACTIVITY: N-(5-phospho-beta-D-riboseyl)-anthranilate = 1-
 CC (2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate.
 CC -!- PATHWAY: Tryptophan biosynthesis; third step.
 CC -!- SIMILARITY: BELONGS TO THE TRPF FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE010345; AA001723.1; -
CC HAMAP; MF00135; -; 1.
CC InterPro; IPR001240; PRAI.
CC Pfam; PF00697; PRAI; 1.
CC Isomerase; Tryptophan biosynthesis; Complete proteome.
CC SEQUENCE 216 AA; 23294 MW; DBAC90CDEAAE1DD8 CRC64;
CC
CC Query Match 37.3%; Score 41; DB 1; Length 216;
CC Best Local Similarity 58.3%; Pred. No. 22;
CC Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
CC
CC QY 5 GVFLKHDWDGS 16
CC 143 GSGGERHWDAS 154
CC
CC RESULT 18
CC KC12_RAT
CC ID KC12_RAT STANDARD; PRT; 414 AA.
CC AC Q62762;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Casein kinase I, gamma 2 isoform (EC 2.7.1.-) (CKI-gamma 2).
CC GN CSNK1G2.
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC OX NCBI_TaxID=10116;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC TISSUE=Testis;
CC RX MEDLINE=95279411; PubMed=7759525;
CC RA Zhai L., Graves P.R., Robinson L.C., Italiano M., Culbertson M.R.,
CC "Casein kinase I gamma subfamily. Molecular cloning, expression, and
CC characterization of three mammalian isoforms and complementation of
CC defects in the Saccharomyces cerevisiae YCK genes.",
CC J. Biol. Chem. 270:12717-12724(1995).
CC RL J. Biol. Chem. 270:12717-12724(1995).
CC CC -1- FUNCTION: Casein kinases are operationally defined by their
CC preferential utilization of acidic proteins such as caseins as
CC substrates. It can phosphorylate a large number of proteins.
CC Participates in Wnt signaling (By similarity).
CC CC -1- SUBUNIT: Monomer (By similarity).
CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC CC -1- TISSUE SPECIFICITY: Testis.
CC CC -1- PTM: AUTOPHOSPHORYLATED.
CC CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CASEIN KINASE I SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U22297; AAC52201.1; -
CC PIR; B56711; B56711.
CC HSP; Q06486; ICKT.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR002290; Ser_thr_kinase.
CC Pfam; PF00069; pkinase; 1.
CC

DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Wnt signaling pathway; Transferase; Serine/threonine-protein kinase;
KW ATP-binding; Phosphorylation; Multigene family.
FT DOMAIN 46 315 PROTEIN KINASE.
FT NP_BIND 52 60 ATP (BY SIMILARITY).
FT BINDING 75 75 ATP (BY SIMILARITY).
FT ACT_SITE 164 164 BY SIMILARITY.
SQ SEQUENCE 414 AA; 47479 MW; 5445A740B4BD576 CRC64;
CC
CC Query Match 37.3%; Score 41; DB 1; Length 414;
CC Best Local Similarity 46.2%; Pred. No. 42;
CC Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
CC
CC QY 6 VFLGLKHDWDGSP 18
CC 321 VFDYEDWAGKPL 333
CC
CC RESULT 19
CC KC12_HUMAN
CC ID KC12_HUMAN STANDARD; PRT; 415 AA.
CC AC P78368; O00704;
CC DT 15-JUL-1998 (Rel. 36, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Casein kinase I, gamma 2 isoform (EC 2.7.1.-) (CKI-gamma 2).
CC GN CSNK1G2.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC OX NCBI_TaxID=9606;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC TISSUE=Testis;
CC RX MEDLINE=98066772; PubMed=9403068;
CC RA Kitabavashi A.N., Kusuda J., Hirai M., Hashimoto K.,
CC "Cloning and chromosomal mapping of human casein kinase I gamma 2
CC (CSNK1G2).",
CC RL Genomics 46:133-137(1997).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
CC Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
CC Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,
CC Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,
CC Liu S., Attix C., Andreise T., Frankheim M., Amico-Keller G.,
CC Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,
CC Kronmiller B., Arellano A., Montgomery M., Ow D., Nolan M., Trong S.,
CC Kobayashi A., Olsen A.S., Carrano A.V.;
CC Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC CC -1- FUNCTION: Casein kinases are operationally defined by their
CC preferential utilization of acidic proteins such as caseins as
CC substrates. It can phosphorylate a large number of proteins.
CC Participates in Wnt signaling (By similarity).
CC CC -1- SUBUNIT: Monomer (By similarity).
CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC CC -1- TISSUE SPECIFICITY: Testis.
CC CC -1- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).
CC CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CASEIN KINASE I SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U89896; AAB88627.1; -
CC

```
DR EMBL; AF001177; AAC00212.1; -.
DR EMBL; AC005306; AAC26983.1; -.
DR HSSP; Q06486; 1CK1.
DR Genew; HGNC:2455; CSNK1G2.
DR MIM; 602214; -.
DR GO; GO:0004681; F:casein kinase I activity; TAS.
DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR000719; Prot_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Wnt signalling pathway; Transferase; Serine/threonine-protein kinase;
KW ATP-binding; Phosphorylation; Multigene family.
KW DOMAIN 46 316
FT NP_BIND 52 60 ATP (BY SIMILARITY).
FT BINDING 75 75 ATP (BY SIMILARITY).
FT ACT_SITE 165 165 BY SIMILARITY.
FT ACT_SITE 165 165 BY SIMILARITY.
SQ SEQUENCE 415 AA; 47457 MW; 036A39148A1DA038 CRC64;

Query Match 37.3%; Score 41; DB 1; Length 415;
Best Local Similarity 46.2%; Pred. NO. 43;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 6 VFGLKHDWDGSPILK 18
DB 322 VFDEYDWDAGKPL 334

RESULT 20
AT18_HUMAN STANDARD; PRT; 1081 AA.
ID AT18_HUMAN
AC Q8TE60;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ADAMTS-18 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 18) (ADAM-TS 18) (ADAM-TS18).
GN ADAMTS18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21856482; PubMed=11867212;
RA Cal S., Obaya A.J., Llamazares M., Garabaya C., Quesada V.,
RA Lopez-Otin C.;
RT "Cloning, expression analysis, and structural characterization of
RT seven novel human ADAMTSs, a family of metalloproteinases with
RT disintegrin and thrombospondin-1 domains.";
RL Gene 283:49-62(2002).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed in fetal lung, liver, and kidney
CC and in adult brain, prostate, submaxillary gland, and endothelium.
CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 4 TSP type-1 domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; AJ311903; CAC83612.1; -.
DR Genew; HGNC:17110; ADAMTS18.
DR MIM; 607512; -.
DR InterPro; IPR002870; Pep_M12B_propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSPI.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tspl-1; 4.
DR SMART; SM00209; TSPI; 4.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; FALSE_NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00214; DISINTEGRIN_2; FALSE_NEG.
DR PROSITE; PS00092; TSPI; 3.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
KW Repeat; Extracellular matrix.
FT SIGNAL 1 47 POTENTIAL.
FT PROPEP 48 284 BY SIMILARITY.
FT CHAIN 285 1081 ADAMTS-18.
FT DOMAIN 285 497 METALLOPROTEASE.
FT DOMAIN 498 577 DISINTEGRIN-LIKE.
FT DOMAIN 588 643 TSP TYPE-1 1.
FT DOMAIN 644 749 CYS-RICH.
FT DOMAIN 750 876 SPACER.
FT DOMAIN 877 931 TSP TYPE-1 2.
FT DOMAIN 933 991 TSP TYPE-1 3.
FT SITE 254 254 CYSTEINE SWITCH (POTENTIAL).
FT METAL 435 435 BY SIMILARITY.
FT ACT_SITE 436 436 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 439 439 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 445 445 ZINC (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 744 744 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 837 837 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 908 908 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1081 AA; 119655 MW; 0438BF645676461E CRC64;

Query Match 37.3%; Score 41; DB 1; Length 1081;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 FGLKHDWDGSPILK 20
DB 441 FGMHDEGNPCRK 454

RESULT 21
LIPB_BUCBP STANDARD; PRT; 203 AA.
ID LIPB_BUCBP
AC Q89AL8;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lipote-protein ligase B (EC 6.-.-) (Lipoate biosynthesis protein
DE B).
GN LIPB OR BBP249.
OS Buchnera aphidicola (subsp. Baizongia pistaciae).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=135842;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426901; PubMed=12522265;
RA Van Ham R.C.H.J., Kamerbeek J., Palacios C., Rausell C., Abascal F.,
RA Bastolla U., Fernandez J.M., Jimenez L., Postigo M., Silva F.J.,
RA Tamames J., Viguera E., Latorre A., Valencia A., Moran F., Moya A.;
RT "Reductive genome evolution in Buchnera aphidicola.";
```

Proc. Natl. Acad. Sci. U.S.A. 100:581-586(2003)).

-1- FUNCTION: Involved in the attachment of lipoyl groups to proteins, by creating an amide linkage that joins the free carboxyl group of lipotic acid to the epsilon-amino group of a specific lysine residue in lipoylated proteins (By similarity).

-1- PATHWAY: Lipocate biosynthesis.

-1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

-1- SIMILARITY: BELONGS TO THE LIPB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF014016; AAC26976.1; -

DR HAMAP: MF_00013; -; 1.

DR PROSITE: PS01313; LIPB; 1.

KW Ligase; Complete proteome.

SQ SEQUENCE 203 AA; 23438 MW; 937AE013BEC225A2 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 203;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 3 TVGVFGLKHD 12
|:|:| | | |
Db 48 TIGVSGTKHD 57

RESULT 22

ID TRPF_HALVO STANDARD; PRT; 221 AA.

AC P52563;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE N-(5'-phosphoribosyl)anthranilate isomerase (EC 5.3.1.24) (PRAI).

GN TRPF.

OS Halobacterium volcanii (Haloflex volcanii).

OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

OC Halobacteriaceae; Haloflex.

OX NCBI_TaxID=2246;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HFD11;

EX MEDLINE=92165748; PubMed=1537810;

RA Lam W.L., Logan S.M., Doolittle W.F.;

RT "Genes for tryptophan biosynthesis in the halophilic archaeobacterium Haloflex volcanii: the trpFEG cluster.";

RL J. Bacteriol. 174:1694-1697(1992).

CC -1- CATALYTIC ACTIVITY: N-(5-phospho-beta-D-ribose)-anthranilate - 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate.

CC -1- PATHWAY: Tryptophan biosynthesis; third step.

CC -1- SIMILARITY: BELONGS TO THE TRPF FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M83788; AAAY3176.1; -

DR HSSP: Q56320; LDL3; -; 1.

DR HAMAP: MF_00135; -; 1.

DR InterPro: IPR001240; PRAI.

DR Pfam: PF00697; PRAI; 1

KW Isomerase; Tryptophan biosynthesis.

SQ SEQUENCE 221 AA; 22634 MW; 396453808C71FBAD CRC64;

Query Match 36.4%; Score 40; DB 1; Length 221;
Best Local Similarity 53.3%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 2 STVGVFGLKHDWDS 16
| | | | | | | |
Db 133 SGAGGTGETHDWDS 147

RESULT 23

ID CYSQ_ECOLI STANDARD; PRT; 246 AA.

AC P22255;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE CysQ protein.

GN CYSQ OR AMTA OR B4214.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / W3110;

EX MEDLINE=91311397; PubMed=1856684;

RA Fabiny J.M., Jayakumar A., Chinault A.C., Barnes E.M. Jr.;

RT "Ammonium transport in Escherichia coli: localization and nucleotide sequence of the amta gene.";

RL J. Gen. Microbiol. 137:983-989(1991).

RN [2]

RP SEQUENCE FROM N.A.

EX MEDLINE=92105007; PubMed=1729235;

RA Newwald A.F., Krishnan B.R., Briken I., Kulakauskas S., Suziedelis K., Tomcsanyi T., Leyh T.S., Berg D.E.;

RT "CysQ, a gene needed for cysteine synthesis in Escherichia coli K-12 only during aerobic growth.";

RL J. Bacteriol. 174:415-425(1992).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

EX MEDLINE=95334362; PubMed=7610040;

RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L., Blattner F.R.;

RT "Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.8 through 100 minutes.";

RL Nucleic Acids Res. 23:2105-2119(1995).

RN [4]

RP SEQUENCE OF 1-98 FROM N.A.

RC STRAIN=K12;

EX MEDLINE=91042419; PubMed=2172762;

RA Liu J., Beacham I.R.;

RT "Transcription and regulation of the cpdB gene in Escherichia coli K12 and Salmonella typhimurium LT2: evidence for modulation of constitutive promoters by cyclic AMP-CRP complex.";

RL Mol. Genet. 222:161-165(1990).

CC -1- FUNCTION: COULD HELP CONTROL THE POOL OF 3'-PHOSPHADENOSIDE 5'-PHOSPHOSULFATE, OR ITS USE IN SULFITE SYNTHESIS.

CC -1- SUBCELLULAR LOCATION: PROBABLE PERIPHERAL MEMBRANE. PROTEIN LOCALIZED ON THE INNER FACE OF THE CYTOPLASMIC MEMBRANE.

CC -1- INDUCTION: STRONGLY REpressed DURING NITROGEN EXCESS.

CC -1- SIMILARITY: Belongs to the inositol monophosphatase family.

CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE AN AMMONIUM TRANSPORT PROTEIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
DR EMBL; M55170; AAA23444.1; -
DR EMBL; M80795; AAA23657.1; -
DR EMBL; U14003; AAA97110.1; -
DR EMBL; AE000492; AAC77171.1; -
DR EMBL; X54008; CAA37953.1; -
DR PIR; S56439; S56439.
DR EcoGene; EG10043; cysQ.
DR InterPro; IPR006240; Bisphos_bact.
DR InterPro; IPR000760; Inositol_P.
DR Pfam; PF00459; inositol_P; 1.
DR ProDom; PD023420; inositol_P; 1.
DR TIGRFAMS; TIGR01331; bisphos_cysQ; 1.
DR PROSITE; PS00629; IMP_1; 1.
DR PROSITE; PS00630; IMP_2; 1.
KW Complete proteome.
FT CONFLICT 193 HV -> QL (IN REF. 3).
SQ SEQUENCE 246 AA; 27171 MW; 13D2DFAA8918E94F CRC64;

Query Match 36.4%; Score 40; DB 1; Length 246;
Best Local Similarity 62.5%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 11 HDWDGSP1 18
DB 221 HDWQGRPL 228

RESULT 24
GBLP_NICPL STANDARD; PRT; 326 AA.
AC P93340;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit-like protein.
OS Nicotiana plumbaginifolia (Leadwort-leaved tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20225456; PubMed=10760577;
RA Kaydanov C., Teves A., Adler K., Manteuffel R.;
RT "Molecular characterization of cDNAs encoding G protein alpha and
RT beta subunits and study of their temporal and spatial expression
RT patterns in Nicotiana plumbaginifolia Viv. ";
RL Biochim. Biophys. Acta 1491:143-160(2000).
CC -1- FUNCTION: MAY HAVE A ROLE IN CELL DIVISION.
CC -1- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y09514; CAA70705.1; -
CC PIR; T16987; T16987.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 7.
CC ProDom; PD000018; WD40; 4.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00678; WD_REPEATS_1; 4.
CC PROSITE; PS50082; WD_REPEATS_2; 6.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Repeat; WD repeat.
CC REPEAT 14 45 WD 1.
CC REPEAT 62 92 WD 2.
CC REPEAT 104 134 WD 3.
CC REPEAT 148 180 WD 4.
CC REPEAT 192 222 WD 5.
CC REPEAT 233 262 WD 6.
CC REPEAT 292 322 WD 7.
CC VARIANT 25 25 A -> P.
CC VARIANT 46 46 I -> L.
CC VARIANT 116 116 V -> A.
CC VARIANT 142 142 T -> I.
CC VARIANT 188 188 L -> A.
CC FT REPEAT 14 45 WD 1.
CC FT REPEAT 62 92 WD 2.
CC FT REPEAT 104 134 WD 3.
CC FT REPEAT 148 180 WD 4.
CC FT REPEAT 192 222 WD 5.
CC FT REPEAT 233 262 WD 6.
CC FT REPEAT 292 322 WD 7.
CC FT VARIANT 25 25 A -> P.
CC FT VARIANT 46 46 I -> L.
CC FT VARIANT 116 116 V -> A.
CC FT VARIANT 142 142 T -> I.
CC FT VARIANT 188 188 L -> A.
```

```
FT REPEAT 148 180 WD 4.
FT REPEAT 192 222 WD 5.
FT REPEAT 233 262 WD 6.
FT REPEAT 292 322 WD 7.
SQ SEQUENCE 326 AA; 92C80F769CAC8100 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 326;
Best Local Similarity 50.0%; Pred. No. 48;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDG 15
DB 73 SSDGMFALSGWDG 86

RESULT 25
GBLP_TOBAC STANDARD; PRT; 326 AA.
ID GBLP_TOBAC
AC P49026;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit-like protein.
GN ARCA.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Bright Yellow 2;
RX MEDLINE=94068562; PubMed=8248221;
RA Ishida S., Takahashi Y., Nagata T.;
RT "Isolation of cDNA of an auxin-regulated gene encoding a G protein
RT beta subunit-like protein from tobacco BY-2 cells. ";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11152-11156(1993).
CC -1- FUNCTION: MAY HAVE A ROLE IN CELL DIVISION.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN ROOTS. ALSO FOUND
CC IN LEAVES, FLOWER BUDS, SHOOT TIPS, STEMS AND YOUNG SEEDLINGS.
CC -1- SIMILARITY: Contains 7 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U03559; BAA04478.1; -
CC PIR; T02340; T02340.
CC InterPro; IPR001680; WD40.
CC Pfam; PF00400; WD40; 7.
CC PRINTS; PR00320; GPROTEINBRPT.
CC ProDom; PD000018; WD40; 4.
CC SMART; SM00320; WD40; 7.
CC PROSITE; PS00678; WD_REPEATS_1; 4.
CC PROSITE; PS50082; WD_REPEATS_2; 6.
CC PROSITE; PS50294; WD_REPEATS_REGION; 1.
CC Repeat; WD repeat.
CC REPEAT 14 45 WD 1.
CC REPEAT 62 92 WD 2.
CC REPEAT 104 134 WD 3.
CC REPEAT 148 180 WD 4.
CC REPEAT 192 222 WD 5.
CC REPEAT 233 262 WD 6.
CC REPEAT 292 322 WD 7.
CC VARIANT 25 25 A -> P.
CC VARIANT 46 46 I -> L.
CC VARIANT 116 116 V -> A.
CC VARIANT 142 142 T -> I.
CC VARIANT 188 188 L -> A.
CC FT REPEAT 14 45 WD 1.
CC FT REPEAT 62 92 WD 2.
CC FT REPEAT 104 134 WD 3.
CC FT REPEAT 148 180 WD 4.
CC FT REPEAT 192 222 WD 5.
CC FT REPEAT 233 262 WD 6.
CC FT REPEAT 292 322 WD 7.
CC FT VARIANT 25 25 A -> P.
CC FT VARIANT 46 46 I -> L.
CC FT VARIANT 116 116 V -> A.
CC FT VARIANT 142 142 T -> I.
CC FT VARIANT 188 188 L -> A.
```

```
FT VARIANT 200 200 P -> T.
FT VARIANT 267 267 S -> T.
FT VARIANT 284 284 S -> A.
FT VARIANT 301 301 S -> G.
FT VARIANT 324 324 D -> G.
SQ SEQUENCE 326 AA; 35945 MW; 680DEDA14AB648BD CRC64;

Query Match 36.4%; Score 40; DB 1; Length 326;
Best Local Similarity 50.08; Pred. No. 48;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDG 15
   I::I::I::I
Db 73 SSDGWFALSGSWDG 86

RESULT 26
YA67_METTH STANDARD; PRT; 377 AA.
AC 027139;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MTH1067.
GN MTH1067.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum M., Pethier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patweli D., Prabakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noellings J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH; functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997);
CC -!- SIMILARITY: STRONG, TO M.JANNASCHII MJ1678 AND A.FULGIDUS AF0028
CC AND AF0181.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000877; AAB85556.1; -
CC PIR; F69008; F69008.
CC KW Hypothetical protein. Complete proteome.
CC SEQUENCE 377 AA; 42021 MW; 24C0E1FA77C7AE4 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 377;
Best Local Similarity 45.5%; Pred. No. 56;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 VGVFGLKHDWDG 14
   I::I::I::I
Db 222 IGVLMKSEWE 232

RESULT 27
PEPD_HAEIN STANDARD; PRT; 483 AA.
ID PEPD_HAEIN
DT 01-NOV-1995 (Rel. 32, Created)
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Proteus.
```

```
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aminoacyl-histidine dipeptidase (EC 3.4.13.3) (Xaa-His dipeptidase)
DE (X-His dipeptidase) (Beta-alanyl-histidine dipeptidase) (Carnosinase)
DE (Peptidase D).
GN PEPD OR HI0675.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: THIS DIPEPTIDASE HAS SPECIFICITY FOR THE UNUSUAL
CC DIPEPTIDE BETA-ALANYL-L-HISTIDINE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of Xaa-l-His dipeptides.
CC -!- COFACTOR: MANGANESE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M25.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32750; AAC22335.1; -
CC PIR; E64085; E64085.
CC DR MEROPS; M25.001; -
CC DR TIGR; HI0675; -
CC DR InterPro; IPR002933; Peptidase_M20.
CC DR InterPro; IPR001160; XHis_dipeptidse.
CC DR Pfam; PF01546; Peptidase_M20; 1.
CC DR PRINTS; PR00934; XHISDIPTASE.
CC KW Hydrolase; Dipeptidase; Metalloprotease; Manganese; Complete proteome.
CC INIT_MET 0 BY SIMILARITY.
CC SEQUENCE 483 AA; 52757 MW; 5591D61750684E4C CRC64;

Query Match 36.4%; Score 40; DB 1; Length 483;
Best Local Similarity 53.3%; Pred. No. 73;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 GVFGKLDWDGSPIL 19
   I::I::I::I
Db 150 GAKGLRHNLQSEIL 164

RESULT 28
UREL_PROMI STANDARD; PRT; 567 AA.
AC P17086.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Urease alpha subunit (EC 3.5.1.5) (Urea amidohydrolase).
DE UREU.
GN Proteus mirabilis.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Proteus.
```

OX NCBI_TaxID=584;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=HI4320;
RX MEDLINE=90078080; PubMed=2687233;
RA Jones B.D., Mobley H.L.T.;
RT "Proteus mirabilis urease: nucleotide sequence determination and
comparison with jack bean urease.";
RL J. Bacteriol. 171:6414-6422(1989).
CC -1- CATALYTIC ACTIVITY: Urea + H(2)O = CO(2) + 2 NH(3).
CC -1- COFACTOR: Binds 2 nickel ions per subunit (By similarity).
CC -1- SUBUNIT: (ALPHA, BETA, GAMMA)3 (BY SIMILARITY).
CC -1- PTM: Lys-217 is carbamylated. The carbamoyl group provides the
ligands for the two nickel ions (By similarity).
CC -1- SIMILARITY: BELONGS TO THE UREASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M31834; AAA25669.1; .
DR PIR: D43719; D43719.
DR HSP: P18314; IFWE.
DR MPROPS; M38.UNW; .
DR InterPro: IPR006680; Amidohydro_1.
DR InterPro: IPR005847; Urease.
DR InterPro: IPR005848; Urease.
DR Pfam: PF01979; Amidohydro_1; 1.
DR Pfam: PF00449; urease; 1.
DR PROSITE: PS00145; UREASE_2; 1.
DR PROSITE: PS01120; UREASE_1; 1.
KW Hydrolase; Metal-binding; Nickel.
FT METAL 134 134 NICKEL 2 (BY SIMILARITY).
FT METAL 136 136 NICKEL 2 (BY SIMILARITY).
FT METAL 217 217 NICKEL 1 AND 2 (BY SIMILARITY).
FT METAL 246 246 NICKEL 1 (BY SIMILARITY).
FT METAL 272 272 NICKEL 1 (BY SIMILARITY).
FT METAL 360 360 NICKEL 2 (BY SIMILARITY).
FT ACT_SITE 320 320 BY SIMILARITY.
SQ SEQUENCE 567 AA; 61012 MW; 67794FD5D495A8E1 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 567;
Best Local Similarity 44.4%; Pred. No. 86;
Matches 8; Conservative 2; Mismatches 6; Indels 2; Gaps 1;

QY 3 TVGVFGK--HDWDGSP 18
| | | | | : | :
Db 210 TAGAIGLKIHEDWGATPM 227

RESULT 29
NADB_ANASP
ID NADB_ANASP STANDARD; PRT; 578 AA.
AC O8YXJ6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE L-aspartate oxidase (EC 1.4.3.16) (LASPO) (Quinolinate synthetase B).
GN NADB OR ALR1217.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,

RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -1- FUNCTION: Catalyzes the oxidation of L-aspartate to
iminoaspartate.
CC -1- CATALYTIC ACTIVITY: L-aspartate + H(2)O + O(2) = oxaloacetate +
NH(3) + H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: NAD biosynthesis; aspartate to NAMN; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FAD-DEPENDENT OXIDOREDUCTASE FAMILY 2.
CC NADB SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP003585; BAB73174.1; .
DR PIR: AF1958; AF1958.
DR InterPro: IPR003953; FAD_bind2.
DR InterPro: IPR005288; NADB.
DR InterPro: IPR004112; Succ_DH_flav_C.
DR Pfam: PF00890; FAD_binding_2; 1.
DR Pfam: PF02910; succ_DH_flav_C; 1.
DR TIGRFAMS: TIGR00551; nadB; 1.
KW Pyridine nucleotide biosynthesis; Oxidoreductase; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 32 46 FAD (AMP PART) (POTENTIAL).
FT ACT_SITE 252 252 BY SIMILARITY.
FT ACT_SITE 271 271 BY SIMILARITY.
SQ SEQUENCE 578 AA; 63172 MW; 0A116B14EAFCBBD8 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 578;
Best Local Similarity 63.6%; Pred. No. 87;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 10 KHDWDGSPILK 20
| | | | | : | :
Db 567 KHQWQSPIMK 577

RESULT 30
YB14_MYCPN
ID YB14_MYCPN STANDARD; PRT; 600 AA.
AC P75448;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative acetyltransferase MPN114 (EC 2.3.1.-) (C09_orf600).
GN MPN114 OR MP040.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelsreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE CARNITINE/CHOLINE ACETYLTRANSFERASE
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AE000005; AAR95688.1; -
 CC PIR: S73366;
 CC InterPro: IPR000542; Carn_acylttransf.
 CC Pfam: PF00755; Carn_acylttransf; 1; FALSE_NEG.
 CC PROSITE: PS00439; ACYLTRANSF_C_1;
 CC PROSITE: PS00440; ACYLTRANSF_C_2; 1;
 CC KW Hypothetical protein; transference; Acyltransferase; Complete proteome.
 CC FT AC SITE 323 323 POTENTIAL.
 CC SQ SEQUENCE 600 AA; 6886 MW; 4791591966AE1E4D CRC64;

Query Match 36.4%; Score 40; DB 1; Length 600;
 Best Local Similarity 46.7%; Pred. No. 91;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 6 VFGLKHDWDGSPILK 20
 DB 387 MFELKWDWFKPLIK 401

RESULT 31

NET1_HUMAN STANDARD; PRT; 604 AA.
 AC O95631;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Netrin-1 precursor.
 GN NTN1 OR NTN1L.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC [1]
 CC NCBI_TaxID=9606;
 RN SEQUENCE FROM N.A., INTERACTION WITH DCC, TISSUE SPECIFICITY, AND
 RP VARIANTS HIS-351 AND GLU-489.
 RC TISSUE=Brain stem, and Liver;
 RX MEDLINE=9913715; PubMed=9950216;
 RA Meyerhardt J.A., Caca K., Eckstrand B.C., Hu G., Lengauer C.,
 RA Banavali S., Look A.T., Fearon E.R.;
 RT "Netrin-1: interaction with deleted in colorectal cancer (DCC) and
 RT alterations in brain tumors and neuroblastomas.";
 RL Cell Growth Differ. 10:35-42(1999).
 CC -1- FUNCTION: Netrins control guidance of CNS commissural axons and
 CC peripheral motor axons.
 CC -1- SUBUNIT: Binds to DCC.
 CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).
 CC -1- TISSUE SPECIFICITY: Widely expressed in normal adult tissues with
 CC highest levels in heart, small intestine, colon, liver and
 CC prostate. Reduced expression in brain tumors and neuroblastomas.
 CC -1- DISEASE: Defects in NTN1 are associated with some forms of
 CC neuroblastomas.
 CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -1- SIMILARITY: Contains 3 laminin EGF-like domains.
 CC -1- SIMILARITY: Contains 1 C345C domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U75586; AAD09221.1; -
 CC HSSP: P02468; 1KLO.
 CC Genew: HGNC:8029; NTN1.
 CC MIM: 601614; -
 CC GO: GO:0007048; P:oncogenesis; TAS.

DR InterPro: IPR006209; EGF like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001886; LamNT.
 DR InterPro: IPR001134; Netrin_C.
 DR Pfam: PF00053; laminin_EGF; 3.
 DR Pfam: PF00055; laminin_EGF; 3.
 DR Pfam: PF01759; NTR; 1.
 DR ProDom: PD002082; Lam_N2; 1.
 DR SMART: SM00643; C345C; 1.
 DR SMART: SM00180; EGF_Lam; 3.
 DR SMART: SM00136; LamNT; 1.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; FALSE_NEG.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 3.
 DR KW Glycoprotein; Extracellular matrix; Signal; Laminin EGF-like domain;
 KW Repeat; Disease mutation.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 604 NETRIN-1.
 FT DOMAIN 25 283 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 285 340 LAMININ EGF-LIKE 1.
 FT DOMAIN 341 403 LAMININ EGF-LIKE 2.
 FT DOMAIN 404 453 LAMININ EGF-LIKE 3.
 FT DOMAIN 454 604 C345C (DOMAIN C).
 FT SITE 530 532 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 119 152 BY SIMILARITY.
 FT DISULFID 285 294 BY SIMILARITY.
 FT DISULFID 287 304 BY SIMILARITY.
 FT DISULFID 306 315 BY SIMILARITY.
 FT DISULFID 318 338 BY SIMILARITY.
 FT DISULFID 341 350 BY SIMILARITY.
 FT DISULFID 343 368 BY SIMILARITY.
 FT DISULFID 371 380 BY SIMILARITY.
 FT DISULFID 383 401 BY SIMILARITY.
 FT DISULFID 404 416 BY SIMILARITY.
 FT DISULFID 406 423 BY SIMILARITY.
 FT DISULFID 425 434 BY SIMILARITY.
 FT DISULFID 437 451 BY SIMILARITY.
 FT DISULFID 476 544 BY SIMILARITY.
 FT DISULFID 491 601 BY SIMILARITY.
 FT CARBOHYD 95 95 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 116 116 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 131 131 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 351 351 R -> H (IN NEUROBLASTOMA).
 FT VARIANT 489 489 K -> E (IN NEUROBLASTOMA).
 FT FTID=VAR_014279.
 FT FTID=VAR_014280.
 SQ SEQUENCE 604 AA; 67734 MW; 332BB62BD16F4691 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 604;
 Best Local Similarity 60.0%; Pred. No. 92;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 HDWDGSPILK 20
 DB 230 HDFSNSPVLO 239

RESULT 32

NET1_MOUSE STANDARD; PRT; 604 AA.
 ID: NET1_MOUSE
 AC O09118; Q60832; Q9QY50;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Netrin-1 precursor.
 DE NTN1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 CC [1]
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.

RC TISSUE=Embryonic brain;
RX MEDLINE=97133208; PubMed=8978605;
RA Serafini T., Colamarino S.A., Leonardo E.D., Wang H., Beddington R.,
RA Skarnes W.C., Tessier-Lavigne M.;
RT "Netrin-1 is required for commissural axon guidance in the developing
RT vertebrate nervous system.";
RL Cell 87:1001-1014(1996).
RN [2]
RN SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=99310665; PubMed=10381568;
RA Puschel A.W.;
RT "Divergent properties of mouse netrins.";
RL Mech. Dev. 83:65-75(1999).
RN [3]
RN SEQUENCE OF 352-398 FROM N.A., AND SUBCELLULAR LOCATION.
RX MEDLINE=95327693; PubMed=7604039;
RA Skarnes W.C., Moss J.E., Hurlley S.M., Beddington R.S.;
RT "Capturing genes encoding membrane and secreted proteins important for
RT mouse development.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:6592-6596(1995).
CC -!- FUNCTION: Netrins control guidance of CNS commissural axons and
CC peripheral motor axons.
CC -!- SUBUNIT: Binds to Dcc (By similarity).
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: In the embryo, widely expressed in the
CC developing nervous system and in mesodermal tissues.
CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC -!- SIMILARITY: Contains 3 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 1 C345C domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U65418; AAC52971.1; .
DR EMBL; AF128865; AAD28602.1; .
DR EMBL; U23505; AAB7938.1; .
DR HSP; P02468; ITL6.
DR MGD; MGI:105088; Ntnl.
DR GO; GO:0005515; F:protein binding activity; IDA.
DR GO; GO:0007411; P:axon guidance; IDA.
DR GO; GO:0030517; P:negative regulation of axon extension; IDA.
DR GO; GO:0040023; P:nuclear positioning; IDA.
DR GO; GO:0045773; P:positive regulation of axon extension; IDA.
DR GO; GO:0030334; P:regulation of cell migration; IDA.
DR InterPro: IPR006209; EGF_like.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001886; LamNT.
DR InterPro: IPR001134; Netrin_C.
DR Pfam; PF00055; laminin_Nterm; 1.
DR Pfam; PF01759; NTR; 1.
DR PRINTS; PR00011; EGFLAMININ.
DR ProDom; PD002082; Lam_N2; 1.
DR SMART; SM00643; C345C; 1.
DR SMART; SM00180; EGF_Lam; 3.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 3.
KW Glycoprotein; Extracellular matrix; Signal; Laminin EGF-like domain;
KW Repeat.
FT SIGNAL. 1 24 POTENTIAL.
FT CHAIN 25 604 NETRIN-1.
FT DOMAIN 25 283 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 285 340 LAMININ EGF-LIKE 1.
FT DOMAIN 341 403 LAMININ EGF-LIKE 2.
FT DOMAIN 404 453 LAMININ EGF-LIKE 3.
FT DOMAIN 454 604 C345C (DOMAIN C).
FT SITE 530 532 CELL ATTACHMENT SITE (POTENTIAL).
FT

FT DISULFID 119 152 BY SIMILARITY.
FT DISULFID 285 294 BY SIMILARITY.
FT DISULFID 287 304 BY SIMILARITY.
FT DISULFID 306 315 BY SIMILARITY.
FT DISULFID 318 338 BY SIMILARITY.
FT DISULFID 341 350 BY SIMILARITY.
FT DISULFID 343 368 BY SIMILARITY.
FT DISULFID 371 380 BY SIMILARITY.
FT DISULFID 383 401 BY SIMILARITY.
FT DISULFID 404 416 BY SIMILARITY.
FT DISULFID 406 423 BY SIMILARITY.
FT DISULFID 425 434 BY SIMILARITY.
FT DISULFID 437 451 BY SIMILARITY.
FT DISULFID 476 544 BY SIMILARITY.
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 604 AA; 67768 MW; 0BE67AE72B837313 CRC64;
Query Match 36.4%; Score 40; DB 1; Length 604;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 11 HDWDGSPILK 20
II: I I I I:
DB 230 HFDNSPVLQ 239

RESULT 33
NET1_CHICK STANDARD; PRT; 606 AA.
AC Q90222;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Netrin-1 precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RX [1]
RC SEQUENCE FROM N.A.
RP STRAIN=White leghorn; TISSUE=Embryonic brain;
RX MEDLINE=94340732; PubMed=8062384;
RA Serafini T., Kennedy T.E., Gallo M.J., Mirzayan C., Jessell T.M.,
RA Tessier-Lavigne M.;
RT "The netrins define a family of axon outgrowth-promoting proteins
RT homologous to C. elegans UNC-6.";
RL Cell 78:409-424(1994).
CC -!- FUNCTION: NETRINS CONTROL GUIDANCE OF CNS COMMISSURAL AXONS AND
CC PERIPHERAL MOTOR AXONS. PROMOTES NEURITE OUTGROWTH FROM
CC COMMISSURAL AXONS BUT ACTS AS A CHEMOREPELLENT FOR TROCHLEAR MOTOR
CC AXONS. THESE EFFECTS ARE MEDIATED BY DISTINCT RECEPTORS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC -!- SIMILARITY: Contains 3 laminin EGF-like domains.
CC -!- SIMILARITY: Contains 1 C345C domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; L34549; AAB60369.1; .
DR PIR; A54665; A54665.
DR HSP; P02468; IKL0.
DR InterPro; IPR006209; EGF_like.
DR

```

DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001886; Laminin_EGF.
DR InterPro: IPR001134; Netrin_C.
DR Pfam: PF00053; laminin_EGF; 3.
DR Pfam: PF00055; laminin_Nterm; 1.
DR Pfam: PF01759; NTR; 1.
DR PRINTS: PR00011; EGFLAMININ.
DR PRODOM: PD002082; Lam_N2; 1.
DR SMART: SM00643; C345C; 1.
DR SMART: SM00180; EGF_Lam; 3.
DR SMART: SM00136; LamNT; 1.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 3.
DR Glycoprotein; Extracellular matrix; Signal; Laminin EGF-like domain;
KW Repeat.
KW SIGNAL
1 25
FT CHAIN 26 606
FT DOMAIN 26 286
FT DOMAIN 287 342
FT DOMAIN 343 405
FT DOMAIN 406 455
FT DOMAIN 456 606
FT SITE 532 534
FT DISULFID 121 154
FT DISULFID 287 296
FT DISULFID 289 306
FT DISULFID 308 317
FT DISULFID 320 330
FT DISULFID 343 352
FT DISULFID 345 370
FT DISULFID 373 382
FT DISULFID 385 403
FT DISULFID 406 418
FT DISULFID 408 425
FT DISULFID 427 436
FT DISULFID 439 453
FT DISULFID 478 546
FT DISULFID 493 603
FT CARBOHYD 97 97
FT CARBOHYD 118 118
FT CARBOHYD 133 133
FT CARBOHYD 419 419
SQ SEQUENCE 606 AA; 68126 MW; 9BF0E3B07A71AE1 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 606;
Best Local Similarity 60.0%; Pred. No. 92;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 HDWDGSPILK 20
Db 232 HFDNSFVLQ 241
11:11111:

RESULT 34
OSTA_XYLFA STANDARD; PRT; 792 AA.
AC QPFA1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Organic solvent tolerance protein precursor.
GN IMP OR OSTA OR XF0837.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
STRAIN=9a5C;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Colauro N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C. de Oliveira R.C. Palmieri D.A. Paris A.,
RA Peixoto B.R., Pereira G.A.G. Pereira H.A. Jr. Pesquero J.B.,
RA Quaggio R.B., Roberto P.G. Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G. Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.P., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Trufi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
RL The genome sequence of the plant pathogen Xylella fastidiosa.;
Nature 406:151-159(2000).
CC -!- FUNCTION: Determines N-hexane tolerance. Involved in outer
CC membrane permeability. Essential for envelope biogenesis. Could be
CC part of a targeting/usher system for outer membrane components (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- SIMILARITY: BELONGS TO THE IMP/OSTA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE003923; AAP83647.1;
CC PIR: B82756; B82756.
CC DR HANAP: MF_01411; -.
CC DR Pfam: PF04453; Osta_C; 1.
CC KW Outer membrane; Signal; Complete proteome.
CC FT SIGNAL 1 22 POTENTIAL
CC FT CHAIN 23 792 ORGANIC SOLVENT TOLERANCE PROTEIN.
CC SQ SEQUENCE 792 AA; 91001 MW; EB97FF8CFD35A422 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 792;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KSTGVFGLKHDW 13
Db 345 QSTGVGTGTGTW 357
:|||||:

RESULT 35
PLSB_XANAC STANDARD; PRT; 885 AA.
ID PLSB_XANAC
AC Q8PES0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycyl-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
GN PLSB OR XAC4270
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.

```

```
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
CC acyl-sn-glycerol 3-phosphate.
CC -!- PATHWAY: DE NOVO PHOSPHOLIPID BIOSYNTHESIS; FIRST STEP. MAY ALSO
CC -!- FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE012079; AM39105.1; -.
DR HAMAP; MF_00393; -.
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
KW phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
KW Complete proteome.
SQ SEQUENCE 885 AA; 98467 MW; FC5942DBAB3B9825 CRC64;
```

```
Query Match 36.4%; Score 40; DB 1; Length 885;
Best Local Similarity 63.6%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

QY 10 KH--DWDGSPF 18
   || |||||
DB 539 KHAPDWDGQPL 549
```

Search completed: October 6, 2003, 07:43:56
Job time : 13.6716 secs

THIS PAGE BLANK (OPTION)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 : Search time 10.7463 Seconds
(without alignments)
178.981 Million cell updates/sec

Title: US-09-765-739A-1
Perfect score: 110
Sequence: 1 KSTGVGFLKHDWDGSPILK 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79	71.8	280	JE0217	28k surface antige
2	72	65.5	276	JE0218	28k surface antige
3	58	52.7	284	I40882	major antigenic pr
4	58	52.7	286	JE0219	28k surface antige
5	53	48.2	278	JE0216	28k surface antige
6	50	45.5	756	T20109	hypothetical prote
7	48	43.6	133	JE0221	28k surface antige
8	47	42.7	396	AS9226	taurapine dehydrog
9	46	41.8	385	A97634	hypothetical prote
10	46	41.8	385	AC2857	conserved hypotet
11	45.5	41.4	231	T32146	hypothetical prote
12	45	40.9	103	F75010	hypothetical prote
13	45	40.9	132	B83834	telchoic acid bios
14	45	40.9	311	E86436	F28K20.3 protein -
15	45	40.9	570	S35137	probable phosphol
16	45	40.9	570	G86777	dihydroxy-acid deh
17	45	40.9	704	DUBPT7	DNA-directed DNA p
18	45	40.9	704	S07512	DNA-directed DNA p
19	45	40.9	819	B87580	conserved hypotet
20	45	40.9	5188	B85547	probable RTX fami
21	45	40.9	5291	F90696	hypothetical prote
22	44	40.0	130	C88102	protein W09G10.6 l
23	44	40.0	160	A75466	2-demethylmenaquin
24	44	40.0	301	D83445	hypothetical prote
25	44	40.0	340	G72596	hypothetical prote
26	44	40.0	348	AE3307	hypothetical prote
27	44	40.0	368	E64142	hypothetical prote
28	44	40.0	410	B59103	hypothetical prote
29	44	40.0	646	T48902	sulfate transporte

30	44	40.0	658	2	T48901	sulfate transporte
31	44	40.0	658	2	T49069	sulfate transporte
32	44	40.0	703	2	T51161	hypothetical prote
33	44	40.0	768	2	E86417	unknown protein, 5
34	44	40.0	770	2	T02318	hypothetical prote
35	44	40.0	886	2	C87031	Cell division prot
36	44	40.0	895	2	T45738	hypothetical prote
37	43	39.1	200	2	T50442	hypothetical prote
38	43	39.1	255	2	S66068	conserved hypotet
39	43	39.1	257	2	D89815	conserved hypotet
40	43	39.1	264	2	G81252	NADH2 dehydrogenas
41	43	39.1	265	2	D71838	nadh oxidoreductas
42	43	39.1	266	2	F64677	NADH2 dehydrogenas
43	43	39.1	306	2	T44684	hypothetical prote
44	43	39.1	334	2	F75539	branched-chain ami
45	43	39.1	386	2	AB2455	glycosyltransferas
46	43	39.1	454	2	S69017	probable membrane
47	43	39.1	492	2	C83592	hypothetical prote
48	43	39.1	686	2	AC0986	probable membrane
49	43	39.1	1186	2	C64588	cag pathogenicity
50	42.5	38.6	84	2	JE0379	NADH2 dehydrogenas
51	42	38.2	141	2	T49411	hypothetical prote
52	42	38.2	150	2	B55209	H transfer determi
53	42	38.2	347	1	QXBO2M	NADH2 dehydrogenas
54	42	38.2	347	2	T11051	NADH2 dehydrogenas
55	42	38.2	404	2	T16272	hypothetical prote
56	42	38.2	413	2	C83225	conserved hypotet
57	42	38.2	469	2	D72722	probable MRSA prot
58	42	38.2	482	2	G75483	probable leucyl am
59	42	38.2	1226	2	T49915	pre-mRNA splicing
60	42	38.2	334	2	JQ1899	genome polyprotein
61	41	37.3	102	2	A53308	hypothetical prote
62	41	37.3	103	2	D71088	hypothetical prote
63	41	37.3	142	2	F83359	hypothetical prote
64	41	37.3	236	2	F83656	hypothetical prote
65	41	37.3	257	2	AB1098	conserved hypotet
66	41	37.3	257	2	AI1460	conserved hypotet
67	41	37.3	259	2	E95071	tributyrin esteras
68	41	37.3	259	2	C97939	tributyrin esteras
69	41	37.3	262	2	AF2117	ABC transporter Af
70	41	37.3	264	2	F72216	hypothetical prote
71	41	37.3	383	2	T15043	fungal elicitor-in
72	41	37.3	383	2	G72359	hypothetical prote
73	41	37.3	396	2	D83520	hypothetical prote
74	41	37.3	405	2	T50717	branched chain ami
75	41	37.3	414	2	B56711	casein kinase I (E
76	41	37.3	430	1	T46965	sulfite dehydrogen
77	41	37.3	557	2	T19241	hypothetical prote
78	41	37.3	561	2	T19242	hypothetical prote
79	41	37.3	612	2	G69797	conserved hypotet
80	41	37.3	725	2	F81845	probable cation-tr
81	41	37.3	746	2	S74219	alpha-galactosidas
82	41	37.3	944	2	S69679	hypothetical prote
83	41	37.3	1209	2	T00373	hypothetical prote
84	41	37.3	4558	2	C82199	RTX toxin RtxA VCI
85	40	36.4	55	2	H95383	protein [imported
86	40	36.4	208	2	T29676	hypothetical prote
87	40	36.4	224	2	C72390	hypothetical prote
88	40	36.4	245	2	T19657	hypothetical prote
89	40	36.4	246	1	S56439	ammonium transport
90	40	36.4	246	2	H91277	ammonium transport
91	40	36.4	246	2	H86118	ammonium transport
92	40	36.4	250	2	T47343	capsid assembly pr
93	40	36.4	290	2	T02300	GTP-binding regula
94	40	36.4	315	2	T47971	seven in absentia-
95	40	36.4	319	2	T36845	probable membrane
96	40	36.4	326	2	T16987	GTP-binding protei
97	40	36.4	326	2	T02340	GTP-binding regula
98	40	36.4	334	2	T16970	GTP-binding protei
99	40	36.4	334	2	G69453	conserved hypotet
100	40	36.4	354	2	D70808	probable regulator

ALIGNMENTS

RESULT 1

JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 71.8%; Score 79; DB 2; Length 280;
Best Local Similarity 73.7%; Pred. No. 6.3e-05; Mismatches 3; Indels 0; Gaps 0;
Matches 14; Conservative 2;

QY 2 STVGVLKHKHWDGSPILK 20

Db 61 TTIGVGLKQWDGSSISK 79

RESULT 2

JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 65.5%; Score 72; DB 2; Length 276;
Best Local Similarity 76.5%; Pred. No. 0.00081; Mismatches 2; Indels 0; Gaps 0;
Matches 13; Conservative 2;

QY 2 STVGVLKHKHWDGSPIL 18

Db 60 TTGVFGLKQNDGSAI 76

RESULT 3

I40882
major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R:van Vliet, A.H.; Jongejans, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1431-1436, 1994
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding the
A:Reference number: I40882; MUID:94178956; PMID:8132352
A:Accession: I40882
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C:Genetics:
A:Gene: map1

Query Match 52.7%; Score 58; DB 2; Length 284;
Best Local Similarity 66.7%; Pred. No. 0.14;

Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KSTGVFGLKHKHWDG 15

Db 61 RDTKAVFGLKKDWDG 75

RESULT 4

JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match 52.7%; Score 58; DB 2; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.14; Mismatches 3; Indels 1; Gaps 0;
Matches 9; Conservative 3;

QY 2 STGVFGLKHKHWD 14

Db 61 TTGVFGLKQDWD 73

RESULT 5

JE0216
28k surface antigen 3 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0216
A:Molecule type: DNA
A:Residues: 1-278 <RED>
A:Cross-references: GB:AF062761

Query Match 48.2%; Score 53; DB 2; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.87; Mismatches 3; Indels 2; Gaps 0;
Matches 8; Conservative 3;

QY 3 TVGVFGLKHKHWDG 15

Db 62 TVALYGLKQDWE 74

RESULT 6

T20109
hypotheoretical protein C50F4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 12-May-2003
C:Accession: T20109
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19225
A:Accession: T20109
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-756 <WIL>
A:Cross-references: EMBL:Z70750; PIDN:CAA94737.1; GSPDB:GN00023; CESP:C50F4.2
A:Experimental source: clone C50F4
C:Genetics:

A;Gene: CESP:C50F4.2
A;Map position: 5
A;Introns: 24/1; 245/3; 315/2; 389/3; 537/3; 568/2; 747/1
C;Superfamily: 6-phosphofructokinase, eukaryotic type; 6-phosphofructokinase 1 homology

Query Match 45.5%; Score 50; DB 2; Length 756;
Best Local Similarity 57.1%; Pred. No. 7.8;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDG 15
| : | | : | | | |
Db 432 SGIQVIGIRKHWG 445
| : | | : | | | |

RESULT 7
JE0221
28k surface antigen 2 - Ehrlichia canis
C;Species: Ehrlichia canis
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
C;Accession: JE0221
R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A;Reference number: JE0216; MUID:98321180; PMID:9647746
A;Accession: JE0221
A;Molecule type: DNA
A;Residues: 1-133 <RED>
A;Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC26722.1; PID:g3327966

Query Match 43.6%; Score 48; DB 2; Length 133;
Best Local Similarity 50.0%; Pred. No. 2.5;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 1 KSTVGVFGLKHDWDGSP1 18
| | | | : | | | |
Db 60 KKTTVYIGLKNWAGDAI 77
| | | | : | | | |

RESULT 8
A59226
taurupine dehydrogenase (EC 1.5.1.23) [validated] - polychaete (Arabella iricolor)
C;Species: Arabella iricolor
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Mar-2001
C;Accession: A59226
R;Kan-no, N.; Sato, M.; Yokoyama, T.; Nagahisa, E.
submitted to the Protein Sequence Database, April 2000
A;Reference number: A59226
A;Accession: A59226
A;Molecule type: protein
A;Residues: 1-396 <KAN>
A;Experimental source: whole body
C;Superfamily: Arabella iricolor taurupine dehydrogenase
C;Keywords: oxidoreductase

Query Match 42.7%; Score 47; DB 2; Length 396;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 11 HDWDGSP1 18
| | | | | : |
Db 221 HDWDGKPV 228
| | | | | : |

RESULT 9
A97634
hypothetical protein AGR_C_4151 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: A97634
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: A97634
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-385 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88026.1; PID:g15157443; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4151
A;Map position: circular chromosome

Query Match 41.8%; Score 46; DB 2; Length 385;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 8; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 4 VGVFGLKHDWDGSPIL 19
: | : | : | | : | |
Db 148 LGLFGM--DWDSTPFL 161
| : | : | : | : |

RESULT 10
AC2857
conserved hypothetical protein Atu2284 [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AC2857
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woc
erage, G.; Gillset, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kar
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AC2857
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-385 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43273.1; PID:g17740760; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2284
A;Map position: circular chromosome

Query Match 41.8%; Score 46; DB 2; Length 385;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 8; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 4 VGVFGLKHDWDGSPIL 19
: | : | : | | : | |
Db 148 LGLFGM--DWDSTPFL 161
| : | : | : | : |

RESULT 11
T32146
hypothetical protein C13A2.11 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32146
R;Rohlfing, T.; Wohldmann, P.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid C13A2.
A;Reference number: Z21126
A;Accession: T32146
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-231 <ROH>
A;Cross-references: EMBL:AF022967; PIDN:AAB69881.1; GSPDB:GN00023; CESP:C13A2.11
A;Experimental source: strain Bristol N2; clone C13A2
C;Genetics:
A;Gene: CESP:C13A2.11
A;Map position: 5
A;Introns: 7/3; 33/3; 71/2; 219/1

A;Experimental source: strain IL1403

C;Genetics:

A;Gene: ilvD

C;Superfamily: dihydroxy-acid dehydratase

C;Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 40.9%; Score 45; DB 2; Length 570;

Best Local Similarity 47.1%; Pred. No. 36;

Matches 8; Conservative 4; Mismatches 3; Indels 2; Gaps 1;

QY 1 KSTVGVLKHDWDGSP 17

|||||

Db 46 KAQGVISM--DWDGNP 60

RESULT 17

DJBPT7

DNA-directed DNA polymerase (EC 2.7.7.7) phage chain - phage T7

N;Alternate names: T7 DNA polymerase

C;Species: phage T7

C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 11-Jun-1999

C;Accession: A00716; S42311

R;Dunn, J.J.; Thompson, K.

submitted to the Nucleic Acid Sequence Database, September 1982

A;Reference number: A94615

A;Accession: A00716

A;Molecule type: DNA

A;Residues: 1-704 <DUN>

R;Dunn, J.J.; Studier, F.W.

J. Mol. Biol. 166, 477-535, 1983

A;Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 ge

A;Reference number: S42283; MUID:83241725; PMID:6864790

A;Accession: S42311

A;Molecule type: DNA

A;Residues: 1-704 <DUW>

A;Cross-references: EMBL:V01146; NID:g431187; PIDN:CAA24412.1; PID:g15591

C;Comment: T7 DNA polymerase is composed of two chains. One is encoded by the phage gene

ach chloroplasts can be substituted for the E. coli subunit and the resultant polymerase

C;Comment: In addition to polymerase activity, T7 DNA polymerase exhibits a 3' to 5' ex

C;Genetics:

A;Gene: 5

A;Map position: 35.94-41.23

C;Superfamily: phage T7 DNA-directed DNA polymerase phage chain

C;Keywords: DNA binding; nucleotidyltransferase

Query Match 40.9%; Score 45; DB 1; Length 704;

Best Local Similarity 87.3%; Pred. No. 45;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 GLKHDWDG 15

|||||

Db 637 GLKHGWDG 644

RESULT 18

S07512

DNA-directed DNA polymerase (EC 2.7.7.7) - phage T3

N;Alternate names: gene 5 protein

C;Species: phage T3

C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 18-Jun-1999

C;Accession: S07512

R;Beck, P.J.; Gonzalez, S.; Ward, C.L.; Molineux, I.J.

J. Mol. Biol. 210, 687-701, 1989

A;Title: Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.

A;Reference number: S07500; MUID:90133923; PMID:2614843

A;Accession: S07512

A;Molecule type: DNA

A;Residues: 1-704 <BEC>

A;Cross-references: EMBL:X17255; NID:g15682; PIDN:CAA35140.1; PID:g15702

C;Genetics:

A;Gene: 5

C;Superfamily: phage T7 DNA-directed DNA polymerase phage chain

C;Keywords: DNA binding; nucleotidyltransferase

Query Match 40.9%; Score 45; DB 2; Length 704;

Best Local Similarity 87.5%; Pred. No. 45;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 GLKHDWDG 15

|||||

Db 637 GLKHGWDG 644

RESULT 19

B87580

conserved hypothetical protein CC2671 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: B87580

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, L.

proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: B87580

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-819 <STO>

A;Cross-references: GB:AE005673; NID:g13424254; PIDN:AAK24638.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2671

Query Match

Best Local Similarity 40.9%; Score 45; DB 2; Length 819;

Matches 10; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 5 GVFGCLKHDW--DGSPIL 19

|||||

Db 170 GVFGAGYDWLPDGSGLF 186

RESULT 20

B85547

probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substr

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C;Accession: B85547

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: B85547

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-5188 <STO>

A;Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: Z0615

Query Match

Best Local Similarity 40.9%; Score 45; DB 2; Length 5188;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 STVGVLKHDWDGS 16

:| || : |||:

Db 4803 TTSGVAAMDYDWDGA 4817

RESULT 21

F90696

hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C:Accession: F90696
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic islands
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: F90696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033965.1; PID:q13360000; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0542

Query Match 40.9%; Score 45; DB 2; Length 5291;
Best Local Similarity 46.7%; Pred. No. 4; Indels 0; Gaps 0;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDG 16
Db 4906 TTSVAAADYDWDGA 4920

RESULT 22
C88102
protein W09G10.6 [imported] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C:Accession: C88102
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology
A:Reference number: A75000; PMID:99059613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999
A:Accession: C88102
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-130 <STO>
A:Cross-references: GB:chr_II; PIDN:AB66114.1; PID:g2315670; GSPDB:GN000020; CESP:W09G10.6
C:Genetics:
A:Gene: W09G10.6
A:Map position: 2

Query Match 40.0%; Score 44; DB 2; Length 130;
Best Local Similarity 60.0%; Pred. No. 10;
Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 1 KSTVGVFGLKHDWDG 15
Db 82 KSTGTGLDGI--DWDG 94

RESULT 23
A75466
2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.1-) DR0859 [similarity] - *Deinococcus*
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: A75466
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A:Reference number: A75250; PMID:20036896; PMID:10567266
A:Accession: A75466
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <WHI>
A:Cross-references: GB:AE001940; GB:AE000513; NID:g6458577; PIDN:AAF10437.1; PID:g645857
A:Experimental source: strain R1
C:Genetics:

A:Gene: DR0859
A:Map position: 1
C:Keywords: methyltransferase

Query Match 40.0%; Score 44; DB 2; Length 160;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 4 VGVFGLKHDWDG 15
Db 79 LGVFGVNGWEG 90

RESULT 24
D83445
hypothetical protein PA1604 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83445
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
A:Reference number: A82950; PMID:20437337; PMID:10984043
A:Accession: D83445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <STO>
A:Cross-references: GB:AE004588; GB:AE004091; NID:g9947563; PIDN:AAG04993.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1604

Query Match 40.0%; Score 44; DB 2; Length 301;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 FGLKHDWDGSPILK 20
Db 153 FGQRDWDADTPILR 166

RESULT 25
G72596
hypothetical protein APE1240 - *Aeropyrum pernix* (strain K1)
C:Species: *Aeropyrum pernix*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72596
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aer*
A:Reference number: A72450; PMID:99310339; PMID:10382966
A:Accession: G72596
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-340 <KAW>
A:Cross-references: DBJ:AP000061; NID:g5104821; PIDN:BAA80229.1; PID:d1044015; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1240

Query Match 40.0%; Score 44; DB 2; Length 340;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 12 DWDGSPILK 20
Db 79 DWDGSPVAR 87

RESULT 26

AE3307
hypothetical protein BMEI0443 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AE3307
R:DeIvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AE3307
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51624.1; PID:g17982351; GSPDB:GN00190
A:Gene: BMEI0443
A:Map position: 1
Query Match 40.0%; Score 44; DB 2; Length 348;
Best Local Similarity 56.2%; Pred. No. 30;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;
QY 4 VGVFGLKHDWDGSPIL 19
:| ||| ||| :|
Db 111 LGFFGL--DWDRTPEL 124

RESULT 27
E64142
hypothetical protein HI0093 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C:Accession: E64142
R:Fielschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J
; Goeyne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7342800
A:Accession: E64142
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-368 <TIGR>
A:Cross-references: GB:U32695; GB:L42023; NID:g1573044; PIDN:AAC21771.1; PID:g1573045; T
A:Note: best homolog was a hypothetical protein from Escherichia coli
C:Superfamily: hypothetical protein HI0093
Query Match 40.0%; Score 44; DB 2; Length 368;
Best Local Similarity 46.2%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 8 GLKHDWDGSPILK 20
|| ||| :|
Db 283 GLSHSQGNELIK 295

RESULT 28
B59103
hypothetical protein pX01-98 - Bacillus anthracis virulence plasmid pX01
C:Species: Bacillus anthracis
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-May-2000
C:Accession: B59103
R:Okinkaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler
J. Bacteriol. 181, 6509-6515, 1999
A:Title: Sequence and organization of pX01, the large Bacillus anthracis plasmid harbori
A:Reference number: A59091; MUID:99445483; PMID:10515943
A:Accession: B59103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <OKI>

A:Cross-references: GB:AF065404; NID:g4894216; PIDN:AAD32402.1; PID:g4894314
A:Experimental source: strain Sterne
C:Genetics:
A:Gene: pX01-98
A:Genome: plasmid

Query Match 40.0%; Score 44; DB 2; Length 410;
Best Local Similarity 53.3%; Pred. No. 36;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 6 VFGFKHDWDGSPILK 20
| | :| | | | |
Db 301 VTGIRYDLDFGPIFK 315

RESULT 29

T48902
sulfate transporter AST12 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T48902
R:Takahashi, H.; Sasakura, N.; Kimura, A.; Watanabe, A.; Saito, K.
Plant Physiol. 121, 686, 1999
A:Title: Identification of two leaf-specific sulfate transporters in Arabidopsis tha
A:Reference number: Z25001
A:Accession: T48902
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-646 <TAK>
A:Cross-references: EMBL:AB012048; PIDN:BAA25175.1
A:Experimental source: ecotype Columbia
C:Genetics:
A:Map position: 3
A:Introns: 116/1; 195/2; 242/3; 264/3; 302/3; 329/3; 372/1; 467/3; 531/2; 560/1; 585/
C:Superfamily: sulfate transport protein

Query Match 40.0%; Score 44; DB 2; Length 646;
Best Local Similarity 46.7%; Pred. No. 60;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGFKHDWDGSPIL 19
| :| | | | | :|
Db 216 GIFGLKHFTDSTDVI 230

RESULT 30

T48901
sulfate transporter ATST1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T48901
R:Sohlberg, L.E.; Sussex, I.M.
Plant Physiol. 113, 1463, 1997
A:Title: Nucleotide sequence of a cDNA (Accession No. U52970) encoding a Cys protein
A:Reference number: Z15626
A:Accession: T48901
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-658 <SOH>
A:Cross-references: EMBL:D89631; PIDN:BAA21657.1
A:Experimental source: Landsberg erecta; tissue type above-ground parts
C:Genetics:
A:Gene: ATST1
C:Superfamily: sulfate transport protein

Query Match 40.0%; Score 44; DB 2; Length 658;
Best Local Similarity 46.7%; Pred. No. 61;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGFKHDWDGSPIL 19
| :| | | | | :|
Db 216 GIFGLKHFTDSTDVI 230

```
RESULT 31
T49069
sulfate transporter (ATST1) - Arabidopsis thaliana
N:Alternate names: protein F4F15.10
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 21-Jul-2000
C:Accession: T49069
R:Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Mewes, H.W.; Rudd, S.; Lemcke, K.;
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25015
A:Cross-references: EMBL:AL049711; GSPDB:GN00061; ATSP:F4F15.10
A:Residues: 1-658 <ALC>
A:Molecule type: DNA
A:Status: preliminary
A:Accession: T49069
A:Map position: 3
A:Gene: ATSP:F4F15.10
A:Introns: 116/1; 185/2; 242/3; 264/3; 302/3; 341/3; 384/1; 479/3; 543/2; 572/1; 597/3
C:Superfamily: sulfate transport protein

Query Match 40.0%; Score 44; DB 2; Length 658;
Best Local Similarity 46.7%; Pred. NO. 61;
Matches 7; Conservative 4; Mismatches 0; Gaps 0;

QY 5 GVFGKHDWDGSPIL 19
I:||||| I: :
DB 216 GIFGLKHFTDSTDVI 230

RESULT 32
T51161
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 02-Sep-2000
C:Accession: T51161
R:Comella, P.; Wu, H.J.; Laudie, M.; Berger, C.; Cooke, R.; Delseny, M.; Grillet, F.
Plant Mol. Biol. 41: 687-700, 1999
A:Title: Fine sequence analysis of 60 kb around the Arabidopsis thaliana AtEm1 locus on
A:Reference number: Z24835; MUID:20108326; PMID:10645728
A:Accession: T51161
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-703 <COM>
A:Cross-references: EMBL:AF049236; PIDN:AAC14417.1
C:Genetics:
A:Map position: 3
A:Introns: 161/1; 230/2; 287/3; 309/3; 347/3; 386/3; 429/1; 524/3; 588/2; 617/1; 642/3
C:Superfamily: sulfate transport protein

Query Match 40.0%; Score 44; DB 2; Length 703;
Best Local Similarity 46.7%; Pred. NO. 65;
Matches 7; Conservative 4; Mismatches 0; Gaps 0;

QY 5 GVFGKHDWDGSPIL 19
I:||||| I: :
DB 261 GIFGLKHFTDSTDVI 275

RESULT 33
E86417
unknown protein, 55790-52851 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86417
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
```

```
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallc
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E86417
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-768 <STO>
A:Cross-references: GB:AE005172; NID:gl0092234; PIDN:AAG12650.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 40.0%; Score 44; DB 2; Length 768;
Best Local Similarity 66.7%; Pred. NO. 72;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 GVFGKHDWD 13
I:|||| I:
DB 659 GLFGIYHDW 667

RESULT 34
T02318
hypothetical protein At2g34300 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein Fl3p17.14
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02318; H84754
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; F
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC Fl3p17 genomic sequence.
A:Reference number: Z14657
A:Accession: T02318
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-770 <ROU>
A:Cross-references: EMBL:AC004481; NID:g3337347; PID:g3337361
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84754
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-770 <STO>
A:Cross-references: GB:AE002093; NID:g3337361; PIDN:AAC27406.1; GSPDB:GN00139
C:Genetics:
A:Gene: Fl3p17.14; At2g34300
A:Map position: 2
A:Introns: 304/3; 354/3; 445/1; 482/1; 540/2; 695/2

Query Match 40.0%; Score 44; DB 2; Length 770;
Best Local Similarity 66.7%; Pred. NO. 72;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 GVFGKHDWD 13
I:|||| I:
DB 661 GLFGIYHDW 669

RESULT 35
C87031
Cell division protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: C87031
R:Colle, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holtr
eam, M.A.; Rutherford, K.M.
```


THIS PAGE BLANK (USPT)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:02 ; Search time 5.67164 Seconds
(without alignments)
165.831 Million cell updates/sec

Title: US-09-765-739A-3

Perfect score: 107

Sequence: 1 NTTGVFGLKQWNGSAISN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	43.0	1004	1 AT1B_ARTSF	P28774 artemia san
2	46	43.0	1034	1 CAPP_SYNY3	P74299 synechocyst
3	45	42.1	160	1 MENG_DEIRA	Q9rw10 deinococcus
4	45	42.1	540	1 YM8M_YEAST	Q03263 saccharomyc
5	43	40.2	86	1 GVPM_HALME	Q02238 halobacteri
6	43	40.2	422	1 YF02_MYCPN	P75285 mycoplasma
7	43	40.2	428	1 YB01_MYCPN	P75285 mycoplasma
8	43	40.2	438	1 YC05_MYCPN	P75571 mycoplasma
9	43	40.2	648	1 VP3_BPPH6	P11129 bacterioph
10	42.5	39.7	1538	1 GLSF_PORPU	P51375 porphyra pu
11	42	39.3	347	1 UTR2_YEAST	P32623 saccharomyc
12	42	39.3	1041	1 ATNA_DROME	P13607 drosophila
13	41	38.3	173	1 YVAR_BACSU	P37506 bacillus su
14	41	38.3	233	1 RR2_CYAPA	P48132 cyanophora
15	41	38.3	282	1 B10B_HELPJ	Q92jk8 helicobacte
16	41	38.3	282	1 B10B_HELPJ	Q25956 helicobacte
17	41	38.3	330	1 Y355_SYNY3	P74436 synechocyst
18	41	38.3	377	1 YA67_METTH	Q27139 methanobact
19	41	38.3	400	1 OAS1_HUMAN	P00573 homo sapien
20	41	38.3	426	1 B2TB_RHOCA	Q52864 rhodobacter
21	41	38.3	741	1 PLO3_MOUSE	Q9r0e1 mus musculu
22	41	38.3	756	1 K6PF_CAEEL	Q27483 caenorhabdi
23	41	38.3	1028	1 SP03_SCHPO	Q9us08 schizosacch
24	40.5	37.9	303	1 NQRF_ALTMA	Q91c13 alteromonas
25	40	37.4	226	1 SP25_CANFA	Q28250 canis famli
26	40	37.4	226	1 SP25_HUMAN	Q15005 homo sapien
27	40	37.4	226	1 SP25_MOUSE	Q9rcy2 mus musculu
28	40	37.4	298	1 HEM3_FUSNN	Q8rfp5 fusobacteri
29	40	37.4	354	1 DHAS_METJA	Q57658 methanococc
30	40	37.4	465	1 YHJA_ECOLI	P37197 escherichia
31	40	37.4	473	1 XYLA_CLOSR	P48790 clostridium
32	40	37.4	543	1 CHG0_MYCPN	P78012 mycoplasma
33	40	37.4	607	1 HRA1_XANCV	P80151 xanthomonas

34	40	37.4	689	1 SYGB_PASMU	P57905 pasteurella
35	40	37.4	792	1 ENTA_XYLFA	Q9pf41 xyella fas
36	40	37.4	942	1 ENV_CAEVG	P31627 caprine art
37	39.5	36.9	303	1 NQRF_ALTHA	Q91c14 alteromonas
38	39.5	36.9	303	1 NQRF_PHOPO	Q91c11 photobacter
39	39.5	36.9	830	1 FARI_YEAST	P21268 saccharomyc
40	39	36.4	161	1 HCRC_THAAR	O33818 thauera aro
41	39	36.4	212	1 IFEL_CAEEL	O45551 caenorhabdi
42	39	36.4	286	1 TYSV_BPT4	P00471 bacterioph
43	39	36.4	308	1 POOB_KLEPN	P27504 klebsiella
44	39	36.4	326	1 GBLP_NICPL	P93340 nicotiana p
45	39	36.4	326	1 GBLP_TOBAC	P49026 nicotiana t
46	39	36.4	371	1 Y028_ARCFU	O30207 archaeoglob
47	39	36.4	388	1 P2X4_HUMAN	O99571 homo sapien
48	39	36.4	418	1 LE21_THEMEA	O9wyc7 thermotoga
49	39	36.4	473	1 SYEL_BRUME	O8yhg4 bruceella me
50	39	36.4	473	1 SYEL_DROME	O890e8 bruceella su
51	39	36.4	475	1 MTHC_DROME	P83119 drosophila
52	39	36.4	527	1 LAC5_TRAVI	Q12717 trametes ve
53	39	36.4	527	1 LAC5_TRAVI	O99056 trametes vi
54	39	36.4	809	1 OSTA_XANCP	O8pce0 xanthomonas
55	39	36.4	813	1 OSTA_XANAC	O8p222 xanthomonas
56	39	36.4	1031	1 ATIA_HYDAT	P35317 hydra atten
57	39	36.4	1033	1 S190_YEAST	P36123 saccharomyc
58	39	36.4	1556	1 GLTS_SYNY3	P55038 synechocyst
59	39	36.4	2282	1 ZAN_RABIT	P57999 oryctolagus
60	39	36.4	2476	1 ZAN_PIG	Q28983 sus scrofa
61	39	36.4	2812	1 ZAN_HUMAN	Q9Y493 homo sapien
62	39	36.4	3066	1 POLG_BCMVN	Q65399 b genome po
63	39	36.4	3080	1 POLG_ZYMVC	P18479 z genome po
64	39	36.4	5376	1 ZAN_MOUSE	O88799 mus musculu
65	38.5	36.0	570	1 C24B_MOUSE	O61093 m cytochrom
66	38.5	36.0	596	1 MKC7_YEAST	P53379 saccharomyc
67	38.5	36.0	795	1 SACE_STRMU	P11701 streptococc
68	38.5	36.0	1113	1 N116_YEAST	Q02630 saccharomyc
69	38	35.5	180	1 RBS_MARPA	O64416 marchantia
70	38	35.5	235	1 RS2_THETN	O8ra21 thermoanaer
71	38	35.5	284	1 YQAK_BACSU	P45908 bacillus su
72	38	35.5	285	1 SYGB_HAEAE	O30836 haemophilus
73	38	35.5	322	1 PTNA_ECOLI	P08186 escherichia
74	38	35.5	341	1 EFG_STRRA	P29541 streptomyc
75	38	35.5	358	1 MBHS_AZOVI	P21950 azotobacter
76	38	35.5	374	1 E13B_HEVBR	P52407 hevea bras
77	38	35.5	484	1 PEPP_ECOLI	P15288 escherichia
78	38	35.5	621	1 Y241_MYCPN	P75441 mycoplasma
79	38	35.5	688	1 SYGB_HAEIN	P43822 haemophilus
80	38	35.5	688	1 SYGB_VIBCH	Q9kwv8 vibrio chol
81	38	35.5	704	1 DPOL_BPT3	P20311 bacterioph
82	38	35.5	704	1 DPOL_BPT7	P00581 bacterioph
83	38	35.5	708	1 EFG1_STRCO	P40173 streptomyc
84	38	35.5	738	1 PLO3_HUMAN	O60568 homo sapien
85	38	35.5	1003	1 ATC_ARTSF	P35316 artemia san
86	38	35.5	1167	1 ITAE_MOUSE	Q06877 mus musculu
87	38	35.5	1206	1 FM14_MOUSE	Q05859 mus musculu
88	38	35.5	1256	1 YH2M_CAEEL	Q27533 caenorhabdi
89	38	35.5	1468	1 FMN1_MOUSE	Q05860 mus musculu
90	38	35.5	3083	1 POLG_ZYMVR	O89330 z genome po
91	38	35.5	3083	1 POLG_ZYMVR	Q36979 z genome po
92	37.5	35.0	137	1 RL16_SPICI	O31162 spiroplasma
93	37.5	35.0	405	1 NQRF_NEIMA	Q9jvq3 neisseria m
94	37.5	35.0	405	1 NQRF_NEIMB	O9k0m8 neisseria m
95	37.5	35.0	407	1 NQRF_PASBU	O9clat6 pasteurella
96	37.5	35.0	500	1 Y039_BORBU	O51068 borrelia bu
97	37.5	35.0	504	1 YM68_YEAST	Q04991 saccharomyc
98	37.5	35.0	570	1 DFA5_ATHAS	Q820c1 anabaena sp
99	37.5	35.0	882	1 SYA_THETH	P74941 thermus the
100	37	34.6	84	1 GVM1_HALNI	P24377 halobacteri

ALIGNMENTS

RESULT 1

AT1B_ARTSF STANDARD: PRT: 1004 AA.

P28774;
01-DEC-1992 (Rel. 24, Created)
01-DEC-1992 (Rel. 24, Last sequence update)
15-SEP-2003 (Rel. 42, Last annotation update)
Sodium/potassium-transporting ATPase alpha chain (EC 3.6.3.9) (Sodium pump) (Na⁺/K⁺ ATPase).
Artemia sanfranciscana (Brine shrimp) (Artemia franciscana).
Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
Artemiidae; Artemia.
NCBI_TaxID=6661;
[1]
SEQUENCE FROM N.A.
MEDLINE=92039032; PubMed=1657719;
MacIas M.T., Martinez J.L., Palmero I., Sastre L.;
"Cloning of a cDNA encoding an Artemia franciscana Na/K ATPase alpha-subunit.";
Gene 105:197-204(1991).
CC -!- FUNCTION: THIS IS THE CATALYTIC COMPONENT OF THE ACTIVE ENZYME,
WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE EXCHANGE OF
NA AND K IONS ACROSS THE PLASMA MEMBRANE. THIS ACTION CREATES THE
ELECTROCHEMICAL GRADIENT OF NA AND K. PROVIDING THE ENERGY FOR
ACTIVE TRANSPORT OF VARIOUS NUTRIENTS.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + Na(+)(In) + K(+)(Out) = ADP +
phosphate + Na(+)(Out) + K(+)(In).
CC -!- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
AND GAMMA.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
ATPases). Subfamily IIC.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; X56650; CAA39972.1; -;
PIR; JH0470; JH0470.
DR DR
HSSP; P04191; 1EUL.
DR DR
InterPro; IPR001757; ATPase_E1-E2.
DR DR
InterPro; IPR006069; Cation_ATPase.
DR DR
InterPro; IPR006068; Cation_ATPase_C.
DR DR
InterPro; IPR004014; Cation_ATPase_N.
DR DR
InterPro; IPR005834; Hydrolase.
DR DR
InterPro; IPR005775; Na/K_ATPase_alpha.
DR DR
Pfam; PF00689; Cation_ATPase_C; 1.
DR DR
Pfam; PF00690; Cation_ATPase_N; 1.
DR DR
Pfam; PF00122; E1-E2_ATPase; 1.
DR DR
Pfam; PF00702; Hydrolase; 1.
DR DR
PRINTS; PR00119; CATATPASE.
DR DR
PRINTS; PR00121; NAKATPASE.
DR DR
TIGRFS; TIGR01106; ATPase-IIC_X-K; 1.
DR DR
TIGRFS; TIGR01494; ATPase_P-type; 4.
DR DR
PROSITE; PS00154; ATPase_E1-E2_1.
DR DR
KW Hydrolase; Sodium/potassium transport; Transmembrane;
KW phosphorylation; ATP-binding. BY SIMILARITY.
DR DR
FT TRANSMEM 76 96
FT TRANSMEM 110 126
FT TRANSMEM 272 294
FT TRANSMEM 301 329
FT TRANSMEM 768 791
FT TRANSMEM 828 855
FT TRANSMEM 897 918
FT TRANSMEM 934 959
FT MOD_RES 357 357
FT BINDING 489 489
FT BINDING 1004 AA; 110699 MW; CE456BCE19A78C7 CRC64;
SQ SEQUENCE 1004 AA; 110699 MW; CE456BCE19A78C7 CRC64;
43.0%; Score 46; DB 1; Length 1004;

Query Match

```

MENG_DEIRA
ID MENG_DEIRA STANDARD; PRT; 160 AA.
AC Q9RW10;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable S-adenosylmethionine:2-demethylmenaquinone methyltransferase
DE (EC 2.1.1.-.-)
GN MENG OR DR0859.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]

SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCBI 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: CONVERTS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK)
CC (By similarity).
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE MENG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: AF001940; AAF10437.1;
CC PIR: A75466; A75466.
CC TIGR: DR0859;
CC HAMAP: MF_00471;
CC InterPro: IPR005493; Methyltransf_6.
CC Pfam: PF03737; Methyltransf_6; 1.
CC Menaquinone biosynthesis; Transferase; Methyltransferase;
CC Complete proteome.
CC SEQUENCE 160 AA; 16860 MW; 54D6F226CAl9EC0E CRC64;

Query Match 42.1%; Score 45; DB 1; Length 160;
Best Local Similarity 46.7%; Pred. No. 3;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 VGVFGLKQWDSAI 18
:||||: | | |
Db 79 LGVFGVENGWGVII 93

RESULT 4
YMBL_YEAST
ID YMBL_YEAST STANDARD; PRT; 540 AA.
AC Q03263;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 59.6 kDa protein in DSK2-CAT8 intergenic region.
GN YMR279C OR YMR021.05C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169872;
RA Bowman S., Church C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY). DHA12 SUBFAMILY.
CC STRONG, TO YEAST ATRI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: Z49704; CAA89777.1;
CC PIR: S54586; S54586.
CC SGD: S0004892; YMR279C.
CC InterPro: IPR007114; MFS.
CC Hypothetical protein; Transport; Transmembrane.
CC TRANSMEM 62 82
CC TRANSMEM 109 129 POTENTIAL.
CC TRANSMEM 132 152 POTENTIAL.
CC TRANSMEM 170 190 POTENTIAL.
CC TRANSMEM 204 224 POTENTIAL.
CC TRANSMEM 233 253 POTENTIAL.
CC TRANSMEM 273 293 POTENTIAL.
CC TRANSMEM 296 316 POTENTIAL.
CC TRANSMEM 335 355 POTENTIAL.
CC TRANSMEM 373 393 POTENTIAL.
CC TRANSMEM 399 419 POTENTIAL.
CC TRANSMEM 430 450 POTENTIAL.
CC TRANSMEM 462 482 POTENTIAL.
CC TRANSMEM 503 523 POTENTIAL.
CC SEQUENCE 540 AA; 59561 MW; 687D06CB0D70AF91 CRC64;

Query Match 42.1%; Score 45; DB 1; Length 540;
Best Local Similarity 36.8%; Pred. No. 10;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 NTTVGVLKQKQWDSAIS 19
| : | : | : | : |
Db 256 NVPTNIHGLSMDWTGSALA 274

RESULT 5
GVPM_HALME STANDARD; PRT; 86 AA.
AC Q02238;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Gas vesicle protein gvpM.
GN GVPM.
OS Halobacterium mediterranei (Haloflex mediterranei).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloflex.
OX NCBI_TaxID=2252;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1411;
RX MEDLINE=93021102; PubMed=1404376;
RA Englert C., Krueger K., Offner S., Pfeifer F.;
RT "Three different but related gene clusters encoding gas vesicles in
RT halophilic archaea."

```

```

J. Mol. Biol. 227:586-592(1992)).
-!- FUNCTION: COULD BE IMPORTANT FOR THE SHAPE DETERMINATION OF THE
    GAS VESICLE.
-!- SIMILARITY: BELONGS TO THE GAS VESICLE PROTEIN TYPE A FAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licensese@isb-sib.ch).
-----
EMBL; X64701; CAA45958.1; -.
PIR; S28126; S28126.
InterPro; IPR000638; Gas_vesicle.
Pfam; PF00741; Gas_vesicle; 1.
ProDom; PD003598; Gas_vesicle; 1.
PROSITE; PS00234; GAS_VESICLE_A_1; 1.
PROSITE; PS00669; GAS_VESICLE_A_2; 1.
Gas vesicle.
SEQUENCE      86 AA;   9355 MW; AB522F20C6F63666 CRC64;
-----
Query Match          40.2%; Score 43; DB 1; Length 86;
Best Local Similarity 40.0%; Pred No. 3,4;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
OY      2 TTGCVGLKQWDGS 16
DB      11:::||||| :
       51 TTTAIVGFNFWDAT 65
-----
RESULT 6
YF02_MYCPN STANDARD; PRT; 422 AA.
AC YF02_MYCPN
PF5285;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
Hypothetical protein MPN502 (P02-orf422V).
MPN502 OR MP341.
Mycoplasmataceae.
Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=2104;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 29342 / M129;
MEDLINE=97105885; PubMed=8948633;
Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
Herrmann R.:
"Complete sequence analysis of the genome of the bacterium Mycoplasma
mycoplaemae".
Nucleic Acids Res. 24:4420-4449(1996).
-!- SIMILARITY: BELONGS TO THE ADHESIN PI FAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licensese@isb-sib.ch).
-----
EMBL; AE000032; AAB95988.1; -.
PIR; S73667; S73667.
Hypothetical protein; Complete proteome.
SEQUENCE      422 AA; 46141 MW; 03BE7D381C7CC298 CRC64;
-----
Query Match          40.2%; Score 43; DB 1; Length 422;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY      1 NTTAVGVFLKL 10

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AE00060; AAB96274.1; -.
 DR PIR; S73952; S73952.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 438 AA; 47925 MW; D6CB89515576AD7E CRC64;

Query Match 40.2%; Score 43; DB 1; Length 438;

Best Local Similarity 70.0%; Pred. No. 18;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTVGVGFLK 10
 |||:|:|
 Db 140 NTTIGAYGLK 149

RESULT 9

ID VP3_BPPH6 STANDARD; PRT; 648 AA.
 AC P11129;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-JAN-1990 (Rel. 13, Last annotation update)
 DE P3 protein.
 GN P3.

OS Bacteriophage phi-6.
 OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
 OX NCBI_TaxID=10879;

RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.
 RX MEDLINE=88160044; PubMed=3347997;

RA Gottlieb P., Metzger S., Romanschuk M., Carton J., Strassman J.,
 RA Bamford D.H., Kalkinen N., Mindich L.;
 RT "Nucleotide sequence of the middle dsRNA segment of bacteriophage phi
 6: placement of the genes of membrane-associated proteins.";

RL Virology 163:183-190(1988).
 CC -!- FUNCTION: P3 PROTEIN IS NECESSARY FOR ADSORPTION ONTO HOST
 CC CELLS.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M17462; AAA68485.1; -.
 DR PIR; C28648; P3BPP6.
 KW Envelope protein.

SQ SEQUENCE 648 AA; 69178 MW; B188DFE02ACC54E3 CRC64;

Query Match 40.2%; Score 43; DB 1; Length 648;

Best Local Similarity 41.2%; Pred. No. 26;
 Matches 7; Conservative 7; Mismatches 1; Indels 2; Gaps 1;

QY 6 VFG--LKQNWGSAIN 20
 :||:|:|:|:|
 Db 181 IFGWYKMDWEGSAVD 197

RESULT 10

ID GLSF_PORPU STANDARD; PRT; 1538 AA.
 AC P51375;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ferredoxin-dependent glutamate synthase (EC 1.4.7.1) (Fd-GOGAT).
 GN GLTB.
 OS Porphyra purpurea.
 OG Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 OX NCBI_TaxID=2787;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Avonport;
 RA Raith M.E., Munnolland J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 genome.";
 RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC -!- CATALYTIC ACTIVITY: 2 L-glutamate + 2 oxidized ferredoxin = L-
 CC glutamine + 2-oxoglutarate + 2 reduced ferredoxin.
 CC -!- COFACTOR: BINDS A 3FE-4S CLUSTER; FAD AND FMN.
 CC -!- PATHWAY: GLUTAMINE SYNTHETASE/GOGAT PATHWAY WHICH IS INVOLVED
 CC IN THE ASSIMILATION OF AMMONIA.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Chloroplast stroma.
 CC -!- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U38804; AAC08261.1; -.
 DR PIR; S73296; S73296.

DR InterPro; IPR002489; DUF14.
 DR InterPro; IPR003009; FMN enzyme.
 DR InterPro; IPR006982; Glu_syn_Central.
 DR InterPro; IPR006981; Glu_syn_NTN.

DR InterPro; IPR002932; Glu_synthase.
 DR Pfam; PF04897; Glu_syn_Central; 1.
 DR Pfam; PF04897; Glu_syn_NTN; 1.

DR Pfam; PF01645; Glu_synthase; 1.
 DR Pfam; PF01493; GXXGG; 1.
 KW Oxidoreductase; Iron-sulfur; 3Fe-4S; Flavoprotein; FAD; FMN;

KW Chloroplast; Glutamate biosynthesis.
 FT DOMAIN 34 384
 FT METAL 1162 1162
 FT METAL 1168 1168
 FT METAL 1173 1173
 FT NE_BIND 1109 1166
 FT SEQUENCE 1538 AA; 168710 MW; 3F647CDB2F5C77CF CRC64;

Query Match 39.7%; Score 42.5; DB 1; Length 1538;
 Best Local Similarity 39.1%; Pred. No. 77;

Matches 9; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY 1 NTTVGVF-----GLKQNWGSAI 18
 ||:|:|:|:|:|
 Db 345 NTEISDFEYXGLQEPWDGPAL 367

RESULT 11
 UTR2_YEAST STANDARD; PRT; 347 AA.
 ID UTR2_YEAST
 AC P32623;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE UTR2 protein (Unknown transcript 2 protein).
 GN UTR2 OR YEL040W OR SYGP-ORF18.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;

RN [1]

CC ACTIVE TRANSPORT OF VARIOUS NUTRIENTS.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + Na(+)(In) + K(+)(Out) = ADP +
CC phosphate + Na(+)(Out) + K(+)(In).
CC -!- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA AND
CC GAMMA.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=1;
CC IsoId=PI3607-1; Sequence=Displayed;
CC Name=2;
CC IsoId=PI3607-2; Sequence=VSP_000417;
CC Name=3;
CC IsoId=PI3607-3; Sequence=VSP_000417, VSP_000418;
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X14476; CAA32638.1; -;
CC EMBL: AF044974; AAC05260.1; -;
CC EMBL: AE003732; AAF55825.2; -;
CC EMBL: AE003732; AAF55827.2; -;
CC EMBL: X17471; CAA35504.1; -;
CC EMBL: U55767; AAB01189.1; -;
CC PIR: S03632; S03632.
CC HSP: P04191; LEUL.
CC FlyBase: FBgn002921; Atp-alpha.
CC InterPro: IPR001757; Atpase_E1-E2.
CC InterPro: IPR006089; Cation_ATPase.
CC InterPro: IPR006068; Cation_ATPase_C.
CC InterPro: IPR004014; Cation_ATPase_N.
CC InterPro: IPR005834; Hydrolase.
CC InterPro: IPR005775; Na/K_ATPase_alph.
CC Pfam: PF00689; Cation_ATPase_C; 1.
CC Pfam: PF00690; Cation_ATPase_N; 1.
CC Pfam: PF00122; E1-E2_ATPase; 1.
CC Pfam: PF00702; Hydrolase; 1.
CC PRINTS: PR00119; CATATPASE.
CC PRINTS: PR00121; NAKATPASE.
CC TIGRFAMS: TIGR01106; Atpase-IIC-X-K; 1.
CC TIGRFAMS: TIGR01494; Atpase-P-type; 5.
CC PROSITE: PS00154; Atpase_E1-E2; 1.
CC Hydrolase; Sodium/potassium transport; Transmembrane;
KW Phosphorylation; ATP-binding; Alternative splicing.
FT TRANSMEM 115 135 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 808 828 POTENTIAL.
FT TRANSMEM 870 890 POTENTIAL.
FT TRANSMEM 935 955 POTENTIAL.
FT TRANSMEM 970 990 POTENTIAL.
FT MOD_RES 394 394 PHOSPHORYLATION (PROBABLE).
FT BINDING 526 526 ATP (BY SIMILARITY).
FT VARSP LIC 1 39 Missing (in isoform 2 and isoform 3).
FT VARSP LIC 844 865 /FTID-VSP_000417.
FT HAAEDIMKRPDPNFKLVNS -> TAEIDMKRPNPF
FT QDKLVNE (in isoform 3).
FT /FTID-VSP_000418.
FT L -> M (IN REF. 1).
FT K -> R (IN REF. 1).
FT MISSING (IN REF. 1).
FT KN -> ED (IN REF. 1).
FT GF -> V (IN REF. 1).
FT 117 118

FT CONFLICT 163 163 I -> V (IN REF. 1).
FT CONFLICT 192 192 E -> G (IN REF. 2).
FT CONFLICT 196 197 LT -> PS (IN REF. 1).
FT CONFLICT 207 208 DV -> VL (IN REF. 1).
FT CONFLICT 211 212 VK -> LE (IN REF. 1).
FT CONFLICT 216 221 RIPADI -> LIPLVY (IN REF. 1).
FT CONFLICT 228 228 N -> D (IN REF. 1).
FT CONFLICT 270 272 GTA -> ALP (IN REF. 1).
FT CONFLICT 290 290 G -> A (IN REF. 1).
FT CONFLICT 299 299 MISSING (IN REF. 1).
FT CONFLICT 402 402 N -> T (IN REF. 4).
FT CONFLICT 488 488 I -> N (IN REF. 4).
FT CONFLICT 811 811 F -> S (IN REF. 1).
FT CONFLICT 843 843 E -> D (IN REF. 1).
SQ SEQUENCE 1041 AA; 115604 MW; B2DD36B2E9029F43 CRC64;

Query Match 39.3%; Score 42; DB 1; Length 1041;
Best Local Similarity 33.3%; Pred. No. 63;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 6 VFGLKQWDGSAISN 20
Db 901 LFGIRKWDKAVND 915
:||::: |||:::

RESULT 13
YYAR_BACSU STANDARD; PRT; 173 AA.
AC P37506;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yyar.
GN YYAR
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RL subtilis chromosome containing the replication origin.";
RN DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haitech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D26185; BAA05205.1; -;
 DR EMBL; Z99124; CAB16111.1; -;
 DR PIR; S65999; S65999.
 DR Subtilist; BG10033; yyar.
 DR InterPro; IPR000182; GCN5acetyltransf.
 DR Pfam; PF00583; Acetyltransf. 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 173 AA; 20393 MW; 0FB09B59F6A056BE CRC64;
 Query Match 38.3%; Score 41; DB 1; Length 173;
 Best Local Similarity 33.3%; Pred. No. 15;
 Matches 6; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
 QY 1 NTTVGVFGLKONWDSAI 18
 Db 74 NNCIGRIKIRSNWNGVAL 91
 RESULT 14
 RR2_CYPARA
 ID RR2_CYPARA STANDARD; PRT; 233 AA.
 AC P46132;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cynelle 30S ribosomal protein S2.
 GN RPS2.
 OS Cyanophora paradoxa.
 OG Cynelle.
 OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
 OX NCBI_TaxID=2762;
 RN [1]
 RC STRAIN=UTEX LB 555 / Pringsheim;
 RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
 RA Bryant D.A.;
 RT "Nucleotide sequence of the cyanelle DNA from *Cyanophora paradoxa*.";
 RL Plant Mol. Biol. Rep. 13:327-332(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=UTEX LB 555 / Pringsheim;
 RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
 RA Farley J.F., Schlutcher W.M., Chung S., Newman-Spallart C.,
 RA Steiner J.M., Jakowitsch J., Bohnert H.J., Bryant D.A.;
 RT The complete sequence of the cyanelle genome of *Cyanophora paradoxa*:
 RT the genetic complexity of a primitive plastid.";
 RL (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
 RL Schwemmler W. (eds.);
 RL Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
 RL (1997).
 CC -----
 CC -1- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U30821; AAA81258.1; -;
 DR PIR; T06915; T06915.
 DR HAMAP; MF_00291; 1.
 DR InterPro; IPR001865; Ribosomal_S2.
 DR InterPro; IPR005706; S2_bact_orf.
 DR Pfam; PF00318; Ribosomal_S2; 1.
 DR PRINTS; PR00395; RIBOSOMALS2.
 DR TIGRFS; TIGR01011; rpsB_bact; 1.
 DR PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
 DR PROSITE; PS00963; RIBOSOMAL_S2_2; 1.
 DR Ribosomal protein; Cynelle.
 SQ SEQUENCE 233 AA; 26198 MW; 06D3BAD7DA9A883F CRC64;
 Query Match 38.3%; Score 41; DB 1; Length 233;
 Best Local Similarity 36.8%; Pred. No. 20;
 Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
 QY 2 TTVGVFGLKONWDSAIN 20
 Db 85 TRCGAFYINRWLGGTLIN 103
 RESULT 15
 BIOB_HELPJ
 ID BIOB_HELPJ STANDARD; PRT; 282 AA.
 AC Q92JK8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).
 GN BIOB OR JHP1298.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
 OC Helicobacteraceae; Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen *Helicobacter pylori*.";
 RL Nature 397:178-180(1999).
 CC -1- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
 CC -1- PATHWAY: Biotin biosynthesis; last step.
 CC -1- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE001553; AAD06876.1; -;
 DR PIR; H71823; H71823.
 DR InterPro; IPR002684; Biotin_synth.
 DR InterPro; IPR006638; Elp3.
 DR Pfam; PF04055; Radical_SAM; 1.
 DR SMART; SM00729; Elp3; 1.
 DR TIGRFS; TIGR00433; bioB; 1.
 KW Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
 FT METAL 17 17 IRON-SULFUR (POTENTIAL).
 FT METAL 21 21 IRON-SULFUR (POTENTIAL).
 FT METAL 24 24 IRON-SULFUR (POTENTIAL).

SQ SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAFLB CRC64;

Query Match 38.3%; Score 41; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 GVFGKQKQND 14
I:||||:|:

Db 157 GIFGLNESWE 166

RESULT 16

BIOSH_HELPY STANDARD; PRT; 282 AA.

AC O25956;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).

GN BIOB OR HP1406.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;

OC Helicobacteraceae; Helicobacter.

OX NCBI_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=26695 / ATCC 700392;

RX MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kervilange A.R., Clayton R.A., Sutton G.G.,

RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,

RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,

RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

RA Venter J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter

RT pylori.";

RL Nature 388:539-547(1997).

CC -I- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.

CC -I- PATHWAY: Biotin biosynthesis; last step.

CC -I- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES

CC FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AE000640; AAD08448.1; -.

DR TIGR; F64695; F64695.

DR TIGR; HP1406; -.

DR InterPro; IPR002684; Biotin_synth.

DR InterPro; IPR006638; Elp3.

DR Pfam; PF04055; Radical_SAM; 1.

DR SMART; SM00729; Elp3; 1.

DR TIGRFAMS; TIGR00433; bioB; 1.

KW Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.

FT METAL 17 17 IRON-SULFUR (POTENTIAL).

FT METAL 21 21 IRON-SULFUR (POTENTIAL).

FT METAL 24 24 IRON-SULFUR (POTENTIAL).

FT METAL 24 24 IRON-SULFUR (POTENTIAL).

SQ SEQUENCE 282 AA; 31475 MW; 8EAD8B601DF151EE CRC64;

Query Match 38.3%; Score 41; DB 1; Length 282;

Best Local Similarity 50.0%; Pred. No. 24;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 GVFGKQKQND 14
I:||||:|:

Db 157 GIFGLNESWE 166

RESULT 17

Y355_SYNY3

AC P74436;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical transport protein sll0355.

GN SLL0355.

OS Synecocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.

OX NCBI_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimpō S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synecocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";

RL DNA Res 3:109-136(1996).

CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -I- SIMILARITY: BELONGS TO THE EMAA TRANSPORTER FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; D90915; BAAL8537.1; -.

DR PIR; S76408; S76408.

DR InterPro; IPR000620; DUF6.

DR Pfam; PF00892; DUF6; 2.

KW Hypothetical protein; Transport; Transmembrane; Complete proteome.

FT TRANSMEM 15 35 POTENTIAL.

FT TRANSMEM 41 61 POTENTIAL.

FT TRANSMEM 72 92 POTENTIAL.

FT TRANSMEM 102 122 POTENTIAL.

FT TRANSMEM 125 145 POTENTIAL.

FT TRANSMEM 175 195 POTENTIAL.

FT TRANSMEM 201 221 POTENTIAL.

FT TRANSMEM 238 258 POTENTIAL.

FT TRANSMEM 264 284 POTENTIAL.

FT TRANSMEM 286 306 POTENTIAL.

SQ SEQUENCE 330 AA; 36092 MW; AB44A99D8BB53DAC CRC64;

Query Match 38.3%; Score 41; DB 1; Length 330;

Best Local Similarity 53.8%; Pred. No. 28;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 8 GLKQNDGSAISN 20

II |||||:

Db 159 GLSINWGSALGS 171

RESULT 18

YA67_METH

ID YA67_METH

AC O27139;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein MTH1067.

GN MTH1067.

OS Methanobacterium thermoaotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria;
 OC Methanobacteriaceae; Methanothermobacter.
 NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwan N., Caruso A., Bush D., Safa H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer C., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RP "Complete genome sequence of Methanobacterium thermoautotrophicum
 RP strain: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ1678 AND A.FULGIDUS AF0028
 CC AND AF0181.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000877; AAB85556.1; -;
 DR PIR: F69008;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 377 AA; 42021 MW; 24C0EE1FA77C7AE4 CRC64;
 Query Match 38.3%; Score 41; DB 1; Length 377;
 Best Local Similarity 33.3%; Pred No. 32;
 Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 4 VGVFGLKQWDCSAI 18
 Db 222 IGVLMKMSWDIVL 236
 ID OAS1_HUMAN
 AC P00973; P04820; P29080; P29081; P78485; P78486; Q16700; Q16701;
 AC Q96J61;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 2'-5'-oligoadenylate synthetase 1 (EC 2.7.7.-) ((2'-5')oligo(A)
 DE synthetase 1) (2-5A synthetase 1) (p46/p42 OAS) (E18/E16).
 GN OAS1 OR OIAs.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM P41 AND P46).
 RX MEDLINE=86081732; PubMed=2416561;
 RA Banach P., Mory J., Revel M., Chebath J.;
 RT "Structure of two forms of the interferon-induced (2'-5') oligo A
 RT synthetase of human cells based on cDNAs and gene sequences.";
 RL EMBO J. 4:2249-2256(1985).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM P41).
 RX MEDLINE=86223945; PubMed=3754863;
 RA Shiojiri S., Fukunaga R., Ichii Y., Sokawa Y.;
 RT "Structure and expression of a cloned cDNA for human (2'-
 RT 5')oligoadenylate synthetase.";
 RL J. Biochem. 99:1455-1464(1986).
 RN [3]

RP SEQUENCE FROM N.A. (ISOFORM P41).
 RX MEDLINE=86108911; PubMed=3753689;
 RA Wathlet M.G., Moutschen S., Cravador A., Dewit L., Defilippi P.,
 RA Huez G.A., Content J.;
 RT "Full-length sequence and expression of the 42 kDa 2-5A synthetase
 RT induced by human interferon.";
 RL FEBS Lett. 196:113-120(1986).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS P48 AND P41).
 RX MEDLINE=91332052; PubMed=1651324;
 RA Ghosh S.K., Kusari J., Bandyopadhyay S.K., Samanta H., Kumar R.,
 RA Sen G.C.;
 RT "Cloning, sequencing, and expression of two murine
 RT 2'-5'-oligoadenylate synthetases. Structure-function relationships.";
 RL J. Biol. Chem. 266:15293-15299(1991).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Muzny D., Arenson A.D., Adams C., Bunac C., Carvelli K., Chang J.,
 RA Chacko J., Chen J., Ding Y., Dugan S., Durbin J., Forcum J.,
 RA Ganesh R., Garcia C., Goodman M., Gorrell J.H., Haywood M.,
 RA Hernandez J., Jackson L., Jin S., Kampal R., Karpthy S., Kovar C.,
 RA Lau S., Leal B., Lee E., Li Y., Lichtarge O., Liu W., Logan O., Lu J.,
 RA Ly T., Marondel I., Martinez C., Merscher S., Montgomery K., Oswal G.,
 RA Perez L., Rashid N.D., Renault B., Rowland K., Savage L.,
 RA Scherer S.E., Shen H., Simon M., Stovall K., Timms K.M., Todd J.,
 RA Vo Q., Williamson A., Worley K.C., Yu W., Kucherlapati R., Nelson D.,
 RA Gibbs R.A.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM P41).
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scaplehorn M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uzdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitling M., Touchman J.W., Green E.D., Dickson M.C.,
 RA Blakesley R.W., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RP SEQUENCE OF 255-364 FROM N.A. (ISOFORM P41).
 RX MEDLINE=8373721; PubMed=634877;
 RA Merlin G., Chebath J., Benach P., Metz R., Revel M.;
 RT "Molecular cloning and sequence of partial cDNA for
 RT interferon-induced (2'-5')oligo(A) synthetase mRNA from human
 RT cells.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4904-4908(1983).
 RN [8]
 RP SEQUENCE OF 231-400 FROM N.A. (ISOFORM P41), AND INDUCTION.
 RC TISSUE=Lymphoblast;
 RX MEDLINE=85284966; PubMed=2411547;
 RA Saunders M.E., Gewert D.R., Tugwell M.E., McMahon M.,
 RA Williams B.R.G.;
 RT "Human 2-5A synthetase: characterization of a novel cDNA and
 RT corresponding gene structure.";
 RL EMBO J. 4:1761-1768(1985).
 RN [9]
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE=88082760; PubMed=3121313;
 RA Wathlet M.G., Claus I.M., Nols C.B., Content J., Huez G.A.;

"New inducers revealed by the promoter sequence analysis of two interferon-activated human genes."; Eur. J. Biochem. 169:313-321(1987).
[10] SEQUENCE OF 1-27 FROM N.A.
TISSUE=Liver;
RX MEDLINE=88142842; PubMed=2830497;
RA Benech P., Vigneron M., Peretz D., Revel M., Chebath J.;
"Interferon-responsive regulatory elements in the promoter of the human 2',5'-oligo(A) synthetase gene."; Mol. Cell. Biol. 7:4498-4504(1987).
[11] SEQUENCE OF 1-27 FROM N.A.
TISSUE=Liver;
RX MEDLINE=88283644; PubMed=2456211;
RA Rutherford M.N., Hannigan G.E., Williams B.R.G.;
"Interferon-induced binding of nuclear factors to promoter elements of the 2-5A synthetase gene."; EMBO J. 7:751-759(1988).
[12] MUTAGENESIS OF CYS-331; PHE-332 AND LYS-333.
RX MEDLINE=98070528; PubMed=9407111;
RA Ghosh A., Sarkar S.N., Guo W., Bandopadhyay S., Sen G.C.;
"Enzymatic activity of 2'-5'-oligoadenylate synthetase is impaired by specific mutations that affect oligomerization of the protein."; J. Biol. Chem. 272:33220-33226(1997).
[13] MUTAGENESIS OF ASP-75 AND ASP-77.
RX MEDLINE=99395121; PubMed=10464285;
RA Sarkar S.N., Ghosh A., Wang H.W., Sung S.S., Sen G.C.;
"The nature of the catalytic domain of 2'-5'-oligoadenylate synthetases."; J. Biol. Chem. 274:25535-25542(1999).
CC -!- FUNCTION: MAY PLAY A ROLE IN MEDIATING RESISTANCE TO VIRUS INFECTION, CONTROL OF CELL GROWTH, DIFFERENTIATION, AND APOPTOSIS.
CC -!- CATALYTIC ACTIVITY: Binds double-stranded RNA and polymerizes ATP into ppp(A2'p5'A)n oligomers, which activate the latent RNase L that, when activated, cleaves single-stranded RNAs.
CC -!- SUBUNIT: Homotetramer.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH DIFFERENT SUBCELLULAR FRACTIONS SUCH AS MITOCHONDRIAL, NUCLEAR, AND ROUGH/SMOOTH MICROSOMAL FRACTIONS.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=3;
CC Name=p46; Synonyms=46 kDa, E18;
CC IsoId=p00973-1; Sequence=Displayed;
CC Name=p41; Synonyms=41 kDa, E16_3-9;
CC IsoId=p00973-2; Sequence=VSP_003738, VSP_003739;
CC Name=p48; Synonyms=9-2;
CC IsoId=p00973-3; Sequence=VSP_003740;
CC -!- INDUCTION: By interferons.
CC -!- SIMILARITY: BELONGS TO THE 2-5A SYNTHETASE FAMILY.
CC -!- CAUTION: Ref.8 sequence differs from that shown due to a frameshift in position 400.
CC -!- CAUTION: REF.4 SEQUENCE WAS ORIGINALLY THOUGHT TO ORIGINATE FROM MOUSE BUT IS A HUMAN SEQUENCE.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X02875; CAA26634.1; -
EMBL; X02874; CAA26633.1; -
EMBL; M11809; AAB59552.1; -
EMBL; M11805; AAB59552.1; JOINED.
EMBL; M11806; AAB59552.1; JOINED.
EMBL; M11807; AAB59552.1; JOINED.
EMBL; M11808; AAB59552.1; JOINED.
EMBL; M11810; AAB59553.1; -

DR EMBL; D00068; BAA00047.1; -
DR EMBL; X04371; CAA51602.1; -
DR EMBL; M63850; AAA39858.1; ALT_INIT.
DR EMBL; M63849; AAA39857.1; ALT_INIT.
DR EMBL; AC004551; -; NOT_ANNOTATED_CDS.
DR EMBL; BC000562; AAH00562.1; -
DR EMBL; X02661; CAA26497.1; ALT_FRAME.
DR EMBL; X06560; CAA29803.1; -
DR EMBL; X07179; CAA30164.1; -
DR EMBL; M18099; AAA59955.1; -
DR EMBL; A14571; CAA01166.1; -
DR EMBL; A14573; CAA01167.1; -
DR PIR; A91013; SYHU16.
DR PIR; B24359; SYHU18.
DR Genew; HGNC:8086; OAS1.
DR MIM; 164350; -
DR GO; GO:0005737; C.cytoplasm; TAS.
DR GO; GO:0006139; P.nucleobase, nucleoside, nucleotide and nucl. . . ; TAS.
DR InterPro; IPR006117; 25A_SYNTH_2.
DR InterPro; IPR006116; 25A_synth_UB.
DR InterPro; IPR001201; PAP_25A_core.
DR PROSITE; PS00832; 25A_SYNTH_1; 1.
DR PROSITE; PS00833; 25A_SYNTH_2; 1.
DR PROSITE; PS0152; 25A_SYNTH_3; 1.
DR RNA-binding; Transferase; Nucleotidyltransferase;
KW Interferon induction; Alternative splicing.

Query Match. 38.3%; Score 41; DB 1; Length 400;
Best Local Similarity 60.0%; Pred. No. 34;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 QNWDGSAISN 20
DB 333 KNWDGSPVSS 342
:|||||:|:

RESULT 20
BZTB_RHOCA STANDARD; PRT; 426 AA.
ID BZTB_RHOCA
AC O52664;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamate/glutamine/aspartate/asparagine transport system permease protein bztb.
DE protein bztb.
GN BZTB
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003 / St. Louis;
RX MEDLINE=96405624; PubMed=8809753;
RA Zheng S., Haseikorn R.;
RT "A glutamate/glutamine/aspartate/asparagine transport operon in Rhodobacter capsulatus."; Mol. Microbiol. 20:1001-1011(1996).
RL -!- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR GLUTAMATE, GLUTAMINE, ASPARTATE AND ASPARAGINE. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC -!- SUBUNIT: BZTB AND BZTC FORM A HETERODIMER WHICH CAN FORM A MEMBRANE COMPLEX WITH A HOMODIMER OF BZTD (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM PERMEASE FAMILY. HISMQ SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; U37407; AAB17887.1; -
 DR PTR; S77606; S77606.
 DR InterPro; IPR000515; BPD_transp.
 DR Pfam; PF00528; BPD_transp; 1.
 DR PROSITE; PS00402; BPD_TRANSPO_INN_MEMBR; 1.
 KW Transport; Amino-acid transport; Transmembrane; Inner membrane.
 FT TRANSMEM 25 45 POTENTIAL.
 FT TRANSMEM 96 116 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 211 231 POTENTIAL.
 FT TRANSMEM 252 272 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT TRANSMEM 340 360 POTENTIAL.
 FT TRANSMEM 396 416 POTENTIAL.
 SQ SEQUENCE 426 AA; 47274 MW; F3277FDD77CE0D8E CRC64;

Query Match 38.3%; Score 41; DB 1; Length 426;
 Best Local Similarity 58.3%; Pred. No. 37;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 2 TTGVFGLKQNW 13
 i : | | | | |
 Db 111 TTIGVRLSQNW 122

RESULT 21

ID PLO3_MOUSE STANDARD; PRT; 741 AA.
 AC O9R0E1; O9CYV9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 precursor
 DE (EC 1.14.11.4) (Lysyl hydroxylase 3) (LH3).
 GN PLOD3.
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RX MEDLINE=99357020; PubMed=10429951;
 RA Ruotsalainen H., Sipila L., Kerkela E., Pospiech H., Myllylae R.;
 RT "Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3,
 RT their phylogenetic analysis and tissue-specific expression in the
 RT mouse.";
 RL Matrix Biol. 18:325-329(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/SVJ;
 RX MEDLINE=21233587; PubMed=11334715;
 RA Ruotsalainen H., Vanhatupa S., Tampio M., Sipila L., Valtavaara M.,
 RA Myllylae R.;
 RT "Complete genomic structure of mouse lysyl hydroxylase 2 and lysyl
 RT hydroxylase 3/collagen glucosyltransferase.";
 RL Matrix Biol. 20:137-146(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- FUNCTION: FORMS HYDROXYLYSINE RESIDUES IN -XAA-LYS-GLY- SEQUENCES
 CC IN COLLAGENS. THESE HYDROXYLYSINES SERVE AS SITES OF ATTACHMENT
 CC FOR CARBOHYDRATE UNITS AND ARE ESSENTIAL FOR THE STABILITY OF THE
 CC INTERMOLECULAR COLLAGEN CROSSLINKS.
 CC -1- CATALYTIC ACTIVITY: Procollagen L-lysine + 2-oxoglutarate + O(2) -
 CC procollagen 5-hydroxy-L-lysine + succinate + CO(2).
 CC -1- COFACTOR: IRON AND ASCORBATE.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE BOUND IN CISTERNAE OF ROUGH
 CC ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the heart, lung, liver and
 CC testis.
 CC -1- SIMILARITY: BELONGS TO THE LYSYL HYDROXYLASE FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF045783; AAD54618.1; -
 DR EMBL; AF014830; AAK00576.1; -
 DR EMBL; AK013195; BAB28704.1; -
 DR MGD; MGI:1347008; Plod3.
 DR InterPro; IPR005123; 2OG-Fell_Oxy.
 DR InterPro; IPR006620; Pro_4_hyd_alph.
 DR InterPro; IPR001006; Procollys_dioxy.
 DR Pfam; PF03171; 2OG-Fell_Oxy; 1.
 DR ProDom; PD011578; Procollys_dioxy; 1.
 DR SMART; SM00702; P4HC; 1.
 DR PROSITE; PS01325; LYS_HYDROXYLASE; 1.
 DR Oxidoreductase; Dioxygenase; Signal; Iron; Vitamin C; Glycoprotein;
 KW Endoplasmic reticulum; Membrane; Polymorphism.
 FT SIGNAL 1 27
 FT CHAIN 28 741
 FT METAL 670 670
 FT METAL 672 672
 FT METAL 722 722
 FT ACT_SITE 732 732
 FT CARBOHYD 66 66
 FT CARBOHYD 286 286
 FT CARBOHYD 551 551
 FT CONFLICT 8 8
 SQ SEQUENCE 741 AA; 84922 MW; DIB79B38639D9F4 CRC64;

Query Match 38.3%; Score 41; DB 1; Length 741;
 Best Local Similarity 42.1%; Pred. No. 64;
 Matches 8; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Oy 1 NTTGVFGLKQNWGSAIS 19
 i : | | | | | :
 Db 66 NYTVRTLGLQSWRGDVA 84

RESULT 22

ID K6PF-CAEEL
 AC Q27483;
 DT 15-JUL-1998 (Rel. 36, Created)


```
RC STRAIN=IAM 12920;
RA MEDLINE=20242140; PubMed=10779868;
RX Kato S., Yumoto I.;
RT "Detection of the Na(+)-translocating NADH-quinone reductase in marine
RI bacteria using a PCR technique.";
RL Can. J. Microbiol. 46:325-332(2000).
CC -!- FUNCTION: NOR COMPLEX CATALYZES THE REDUCTION OF UBIQUINONE-1 TO
CC UBIQUINOL BY TWO SUCCESSIVE REACTIONS, COUPLED WITH THE TRANSPORT
CC OF NA(+) IONS FROM THE CYTOPLASM TO THE PERIPLASM. THE FIRST STEP
CC IS CATALYZED BY NORF, WHICH ACCEPTS ELECTRONS FROM NADH AND
CC REDUCES UBIQUINONE-1 TO UBISEMIOQUINONE BY A ONE-ELECTRON TRANSFER
CC PATHWAY.
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone + Na(+)(In) = NAD(+) +
CC ubiquinol + Na(+)(Out).
CC -!- COFACTOR: FAD AND A 2FE-2S CLUSTER (BY SIMILARITY).
CC -!- SUBUNIT: COMPOSED OF SIX SUBUNITS; NQRA, NQRB, NQRC, NQRD, NQRE
CC AND NQRF (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Inner membrane (Potential).
CC -!- SIMILARITY: BELONGS TO THE NQRF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB024721; BAA83758.1; -
CC HAMAP; MF_00430; -; 1.
CC InterPro; IPR006058; 2Fe2S_ferredoxin.
CC InterPro; IPR001834; Cyt_B5_reductase.
CC InterPro; IPR001041; Ferredoxin.
CC InterPro; IPR001433; Oxred_FAD/NAD(P).
CC Pfam; PF00970; FAD_binding_6; 1.
CC Pfam; PF00111; fer2; 1.
CC Pfam; PF00175; NAD_binding_1; 1.
CC PROSITE; PS00197; 2FE2S_FERREDOXIN; PARTIAL.
CC Oxidoreductase; NAD; Ubiquinone; Transport; Sodium transport;
CC Flavoproteins; FAD; Metal-binding; Iron-sulfur; Iron; 2Fe-2S;
CC Inner membrane
CC NON_TER 1
CC DOMAIN <1 45 FERREDOXIN.
CC DOMAIN 201 >303 CATALYTIC.
CC METAL 4 4 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 7 7 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 39 39 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC NON_TER 303
CC SEQUENCE 303 AA; 34505 MW; 56CD2E5CB5931131 CRC64;

Query Match 37.9%; Score 40.5; DB 1; Length 303;
Best Local Similarity 60.0%; Pred. No. 31;
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Qy 6 VFGLKQNDGSAISN 20
Db 55 VFGVKK-WDCEVISN 68

RESULT 25
SP25_CANFA STANDARD; PRT; 226 AA.
ID SP25_CANFA
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microsomal signal peptidase 25 kDa subunit (EC 3.4.-.-) (Spase 25: kDa
DE subunit) (SPC25).
GN SPC25.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
```

```
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=95014327; PubMed=7929230;
RA Greenburg G., Blobel G.;
RT "cDNA-derived primary structure of the 25-kDa subunit of canine
RI microsomal signal peptidase complex.";
RL J. Biol. Chem. 269:25354-25358(1994).
CC [2]
CC TOPOLOGY.
CC MEDLINE=96216505; PubMed=8632014;
CC Kallies K.-U., Hartmann E.;
CC "Membrane topology of the 12- and the 25-kDa subunits of the mammalian
RT signal peptidase complex.";
CC J. Biol. Chem. 271:3925-3929(1996).
CC -!- FUNCTION: MICROSMAL SIGNAL PEPTIDASE IS A MEMBRAN-BOUND
CC ENDOPEPTIDINASE THAT REMOVES SIGNAL PEPTIDES FROM NASCENT
CC PROTEINS AS THEY ARE TRANSLOCATED INTO THE LUMEN OF THE
CC ENDOPLASMIC RETICULUM.
CC -!- SUBUNIT: COMPLEX OF FIVE DIFFERENT PROTEINS: SPC25, SPC22/23,
CC SPC21, SPC18, AND SPC12.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE SPC25 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U12687; AAA21254.1; -
CC PIR; A55012; A55012
CC Hydrolase; Microsome; Endoplasmic reticulum; Transmembrane; Protease.
CC DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 87 107 POTENTIAL.
CC DOMAIN 108 111 LUMENAL (POTENTIAL).
CC TRANSMEM 112 132 POTENTIAL.
CC DOMAIN 133 226 CYTOPLASMIC (POTENTIAL).
CC SEQUENCE 226 AA; 24941 MW; 85830E5484593176 CRC64;

Query Match 37.4%; Score 40; DB 1; Length 226;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 13 WDGSAISN 20
Db 51 WDGSVAVKN 58

RESULT 26
SP25_HUMAN STANDARD; PRT; 226 AA.
ID SP25_HUMAN
AC Q15005; Q15507; Q96HU9.
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Microsomal signal peptidase 25 kDa subunit (EC 3.4.-.-) (Spase 25 kDa
DE subunit) (SPC25).
GN SPC25 OR KIAA0102.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 2-226 FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=95308325; PubMed=7788527;
RA Nagase T., Miyajima K., Tanaka A., Sazuka T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawarabayashi Y., Kotani H., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. III.
The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
```

RT analysis of cDNA clones from human cell line KG-1.";
 RL DNA Res. 2:37-43(1995).
 RN [2]
 RP SEQUENCE OF 1-37 FROM N.A.
 RA Hartmann E.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 32-226 FROM N.A.
 RX TISSUE=Skin;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Bosak S.A., McWan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Touchman J.W., Green E.D., Dickson M.C.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: MICROSOMAL SIGNAL PEPTIDASE IS A MEMBRANE-BOUND
 CC ENDOPROTEINASE THAT REMOVES SIGNAL PEPTIDES FROM NASCENT
 CC PROTEINS AS THEY ARE TRANSLOCATED INTO THE LUMEN OF THE
 CC ENDOPLASMIC RETICULUM.
 CC -!- SUBUNIT: COMPLEX OF FIVE DIFFERENT PROTEINS: SPC25, SPC22/23,
 CC SPC21, SPC18, AND SPC12.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE SPC25 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D14658; BAA03492.1; ALT_INIT.
 DR EMBL; L38950; AAA60992.1; -;
 DR EMBL; BC008063; AAA08063.1; ALT_INIT.
 KW Hydrolase; Microsome; Endoplasmic reticulum; Transmembrane; Protease.
 FT DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 87 107 POTENTIAL.
 FT DOMAIN 108 111 LUMENAL (POTENTIAL).
 FT TRANSMEM 112 132 POTENTIAL.
 FT DOMAIN 133 226 CYTOPLASMIC (POTENTIAL).
 FT CONFLICT 32 33 RS -> HE (IN REF. 3).
 SQ SEQUENCE 226 AA; 25003 MW; C1EDF687E9F7A57A CRC64;
 Query Match 37.4%; Score 40; DB 1; Length 226;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 13 WDGSATSN 20
 |||||:|
 Db 51 WDGSAYKN 58
 RESULT 27
 SP25_MOUSE STANDARD; PRT; 226 AA.
 ID SP25_MOUSE
 AC Q9CYN2; Q921V8; Q9CXK1;
 DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Microsomal signal peptidase 25 kDa subunit (EC 3.4.-.-) (SPase 25 kDa
 DE subunit) (SPC25).
 GN SPC25.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C57BL/6J; TISSUE=Embryo;
 RA MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Tovo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [2]
 RP SEQUENCE OF 60-226 FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramo R.D., Mullaly S.J.,
 RA Bosak S.A., McWan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Touchman J.W., Green E.D., Dickson M.C.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: MICROSOMAL SIGNAL PEPTIDASE IS A MEMBRANE-BOUND
 CC ENDOPROTEINASE THAT REMOVES SIGNAL PEPTIDES FROM NASCENT
 CC PROTEINS AS THEY ARE TRANSLOCATED INTO THE LUMEN OF THE
 CC ENDOPLASMIC RETICULUM.
 CC -!- SUBUNIT: COMPLEX OF FIVE DIFFERENT PROTEINS: SPC25, SPC22/23,
 CC SPC21, SPC18, AND SPC12.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE SPC25 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D14658; BAA03492.1; ALT_INIT.
 DR EMBL; L38950; AAA60992.1; -;
 DR EMBL; BC008063; AAA08063.1; ALT_INIT.
 KW Hydrolase; Microsome; Endoplasmic reticulum; Transmembrane; Protease.
 FT DOMAIN 1 86 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 87 107 POTENTIAL.
 FT DOMAIN 108 111 LUMENAL (POTENTIAL).
 FT TRANSMEM 112 132 POTENTIAL.
 FT DOMAIN 133 226 CYTOPLASMIC (POTENTIAL).
 FT CONFLICT 32 33 RS -> HE (IN REF. 3).
 SQ SEQUENCE 226 AA; 25003 MW; C1EDF687E9F7A57A CRC64;
 Query Match 37.4%; Score 40; DB 1; Length 226;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 13 WDGSATSN 20
 |||||:|
 Db 51 WDGSAYKN 58
 RESULT 27
 SP25_MOUSE STANDARD; PRT; 226 AA.
 ID SP25_MOUSE
 AC Q9CYN2; Q921V8; Q9CXK1;
 DT 28-FEB-2003 (Rel. 41, Created)

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AK014310; BAB29262.1; -;
 DR EMBL; AK01504; BAB30777.1; -;
 DR EMBL; BC010547; AAH10547.1; ALT_INIT.
 DR MGD; MGI:1913874; 5730406115Rik.
 KW Hydrolase; Microsome; Endoplasmic reticulum; Transmembrane; Protease.
 FT TRANSMEM 1 86 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 87 107 POTENTIAL.
 FT TRANSMEM 108 111 LUMENAL (POTENTIAL).
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 133 226 CYTOPLASMIC (POTENTIAL).
 FT CONFLICT 209 209 E -> D (IN REF. 1; BAB29262).
 SQ SEQUENCE 226 AA; 24977 MW; F8516C51FFED4DF9 CRC64;
 Query Match 37.4%; Score 40; DB 1; Length 226;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 13 WDGSAIN 20
 DB 51 WDGSAYKN 58
 RESULT 28
 ID HEM3_FUSNN STANDARD; PRT; 298 AA.
 AC Q8RFP5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Porphobilinogen deaminase (EC 4.3.1.8) (PBG) (Hydroxymethylbilane
 synthase) (HMB) (Pre-uroporphyrinogen synthase).
 GN HEMC OR FN0645.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21896324; PubMed=11899109;
 EX Kapatali V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
 Larsen N., D'Souza M., Walunas T., Pusch G., Haseikorn R.,
 Ronstein M., Kyripides N., Overbeek R.;
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184(2005-2018(2002)).
 CC -!- FUNCTION: Tetrapolymerization of the monopyrrole PBG into the
 hydroxymethylbilane preuroporphyrinogen in several discrete steps.
 CC -!- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =
 CC -!- COFACTOR: Covalently binds a dipyrromethane cofactor to which the
 CC -!- porphobilinogen subunits are added (By similarity).
 CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; fourth step.
 CC -!- SUBUNIT: Monomer (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE HMB5 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE010575; AAL94841.1; -;
 DR HAMAP; MF_00260; -; 1.
 DR InterPro; IPR000860; Porphobil_deam.
 DR Pfam; PF01379; Porphobil_deam; 1.
 DR Pfam; PF03900; Porphobil_deamC; 1.
 DR ProDom; PD002745; Porphobil_deam; 1.
 DR TIGRFAMs; TIGR00212; hemc; 1.

DR PROSITE; PS00533; PORPHOBILINOGEN_DEAM; 1.
 KW Porphyrin biosynthesis; Lyase; Complete proteome.
 FT BINDING 242 242 PYRROMETHANE COFACTOR (BY SIMILARITY).
 SQ SEQUENCE 298 AA; 33098 MW; 13908A7D0AA56984 CRC64;
 Query Match 37.4%; Score 40; DB 1; Length 298;
 Best Local Similarity 52.9%; Pred. No. 37;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 3 TVGVFLKONWDGSAIS 19
 DB 42 TSGDKLKSNNWNSDIS 58
 RESULT 29
 ID DHAS_METJA STANDARD; PRT; 354 AA.
 AC Q57658;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) (ASA
 dehydrogenase) (ASADH).
 GN ASD OR MJ0205.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 3661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL -!- FUNCTION: THIS ENZYME CATALYZES THE SECOND STEP IN THE COMMON
 METABOLIC PATHWAY TO SYNTHESIZE THR AND MET FROM ASPARTIC ACID.
 CC -!- CATALYTIC ACTIVITY: L-aspartate-4-semialdehyde + phosphate +
 CC NADP(+) = L-4-aspartyl phosphate + NADPH.
 CC -!- PATHWAY: METHIONINE BIOSYNTHESIS, THREONINE BIOSYNTHESIS.
 CC -!- SIMILARITY: Belongs to the aspartate-semialdehyde dehydrogenase
 CC family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U67476; AAB98189.1; -;
 DR PIR; F64325; F64325.
 DR TIGR; MJ0205; -;
 DR InterPro; IPR005676; Asp_ADH_f.
 DR InterPro; IPR000319; Asp_semi醛_dh.
 DR InterPro; IPR000534; Semi醛_dh.
 DR Pfam; PF01118; Semi醛_dh_dhc; 1.
 DR Pfam; PF02774; Semi醛_dh_dhc; 1.
 DR TIGRFAMs; TIGR00978; asd_EA; 1.
 DR PROSITE; PS01103; ASD; 1.
 DR Oxidoreductase; NADP; Methionine biosynthesis; Threonine biosynthesis;
 KW Amino-acid biosynthesis; Complete proteome.
 FT ACT_SITE 157 157 BY SIMILARITY.
 SQ SEQUENCE 354 AA; 39439 MW; 5C412CA5C39B166E CRC64;

Query Match 37.4%; Score 40; DB 1; Length 354;
Best Local Similarity 54.5%; Pred. No. 44;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 10 KQWDGSAISN 20
| : | | : | : |
Db 144 KRWDGAIITN 154

```
RESULT 30
YHJA_ECOLI STANDARD; PRT; 465 AA.
AC P37197;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable cytochrome C peroxidase (EC 1.1.1.5).
GN YHJA OR B3518.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- CATALYTIC ACTIVITY: 2 ferrocyclochrome c + H(2)O(2) - 2
CC ferrocyclochrome c + 2 H(2)O.
CC -!- PTM: BINDS 3 HEMES (POTENTIAL).
CC -!- SIMILARITY: HIGH, TO P.AERUGINOSA CYTOCHROME C551 PEROXIDASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00039; AAB18494.1; -.
DR EMBL; AF000428; AAC76543.1; -.
DR PIR; S47738; S47738.
DR HSSP; P14532; LEB7.
DR EcoGene; EG12244; yhjA.
DR InterPro; IPR003088; Cyt_C1.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR004852; CytCP_Maug.
DR Pfam; PF03150; CCP_Maug; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 3.
KW Hypothetical protein; Oxidoreductase; Peroxidase; Heme;
KW Electron transport; Complete proteome.
FT BINDING 59 59 HEME 1 (COVALENT) (BY SIMILARITY).
FT BINDING 62 62 HEME 1 (COVALENT) (BY SIMILARITY).
FT METAL 63 63 IRON (HEME 1 PROXIMAL LIGAND)
(BY SIMILARITY).
FT BINDING 207 207 HEME 2 (COVALENT) (BY SIMILARITY).
FT BINDING 210 210 HEME 2 (COVALENT) (BY SIMILARITY).
FT METAL 211 211 IRON (HEME 2 PROXIMAL LIGAND)
(BY SIMILARITY).
FT BINDING 351 351 HEME 3 (COVALENT) (BY SIMILARITY).
FT BINDING 354 354 HEME 3 (COVALENT) (BY SIMILARITY).
FT METAL 355 355 IRON (HEME 3 PROXIMAL LIGAND) (BY
SIMILARITY).
FT METAL 415 415 IRON (HEME 2 DISTAL LIGAND) (BY
SIMILARITY).
FT METAL 429 429 IRON (HEME 3 DISTAL LIGAND) (BY
SIMILARITY).
FT -SEQUENCE 465 AA; 51570 MW; 9F494A698949E6DA CRC64;
```

Query Match 37.4%; Score 40; DB 1; Length 465;
Best Local Similarity 58.3%; Pred. No. 58;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 VFGLKQNDGSA 17
| : | | | |
Db 243 VENVEQFWDGRA 254

```
RESULT 31
XYLA_CLOSR STANDARD; PRT; 473 AA.
AC P48790;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Xylosidase/arabinosidase [includes: Beta-xylosidase (EC 3.2.1.37)
DE (1,4-beta-D-xylan xylohydrolase) (Xylan 1,4-beta-xylosidase); Alpha-L-
DE arabinofuranosidase (EC 3.2.1.55) (Arabinosidase)].
GN XYLA.
OS Clostridium stercorarium.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1510;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=93214114; PubMed=77633495;
RA Sakka K., Yoshikawa K., Kojima Y., Karita S., Ohmiya K., Shinada K.;
RT "Nucleotide sequence of the Clostridium stercorarium xyla gene
RT encoding a bifunctional protein with beta-D-xylosidase and alpha-L-
RT arabinofuranosidase activities, and properties of the translated
RT product."
RL Biosci. Biotechnol. Biochem. 57:268-272(1993).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-xylans so as to
CC remove successive D-xylose residues from the non-reducing termini.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing alpha-L-
CC arabinofuranoside residues in alpha-L-arabinosides.
CC -!- SUBUNIT: Homotetramer.
CC -!- SIMILARITY: BELONGS TO FAMILY 43 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13268; BAA02527.1; -.
DR PIR; J01936; J01936.
KW Xylan degradation; Hydrolase; Glycosidase; Multifunctional enzyme.
SQ SEQUENCE 473 AA; 53341 MW; CDA34CE9DEBB2399 CRC64;
-----
Query Match 37.4%; Score 40; DB 1; Length 473;
Best Local Similarity 42.9%; Pred. No. 59;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGKQNDGSAI 18
| | | : | : | : |
Db 413 GVFEVKTWSNGEVL 426

RESULT 32
CH60_MYCPN STANDARD; PRT; 543 AA.
ID CH60_MYCPN
AC P78012;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
GN GROEL OR GROEL OR MOPA OR MPN573 OR MP269.
OS Mycoplasma pneumoniae.
```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AE006151; AAK03186.1; -;
 DR HAMAP; MF_00255; -; 1.
 DR InterPro; IPR002311; trna_synt_2f.
 DR InterPro; IPR006194; trna_synt_gly.
 DR Pfam; PF02092; trna_synt_2f; 1.
 DR PRINTS; PR01045; TRNASYNTHGB.
 DR TIGRFAMs; TIGR00211; glyS; 1.
 DR PROSITE; PS00861; AA_TRNA_LIGASE_IL_GLYAB; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; Atp-binding;
 KW Complete proteome.
 SQ SEQUENCE 689 AA; 76019 MW; D67A980B5143B21E CRC64;

Query Match 37.4%; Score 40; DB 1; Length 689;
 Best Local Similarity 47.1%; Pred. No. 87;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 NTTVGVFLKQNWGSA 17
 : : | : | : |
 Db 459 DTLTGIFGIQPKGSA 475

RESULT 35
 OSTA_XYLFA
 ID OSTA_XYLFA STANDARD; PRT; 792 AA.
 AC Q9PF41;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Organic solvent tolerance protein precursor.
 GN IMP OR OSTA OR XF0837.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."
 RL Nature 406:151-159(2000).
 CC -!- FUNCTION: Determines N-hexane tolerance. Involved in outer

CC membrane permeability. Essential for envelope biogenesis. Could be
 CC part of a targeting/usher system for outer membrane components (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE IMP/OSTA FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: AE003923; AAF83647.1; -;
 DR PIR: B82756; B82756.
 DR HAMAP; MF_01411; -; 1.
 DR Pfam; PF04453; OstaA_C; 1.
 KW Outer membrane; Signal; Complete proteome.
 KW SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 792 ORGANIC SOLVENT TOLERANCE PROTEIN.
 SQ SEQUENCE 792 AA; 91001 MW; EB97FF8CFD35A422 CRC64;

Query Match 37.4%; Score 40; DB 1; Length 792;
 Best Local Similarity 31.6%; Pred. No. 1e+02;
 Matches 6; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 2 TTVGVFLKQNWGSAISN 20
 : : | : | : |
 Db 346 STGVGVGTGETWTAGLMAD 364

Search completed: October 6, 2003, 07:44:04
 Job time : 8.67164 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 10.7463 Seconds
(without alignments)
178.981 Million cell updates/sec

Title: US-09-765-739A-3

Perfect score: 107

Sequence: 1 NTTVGVFLKQNDGSAISN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	276	JE0218	28k surface antige
2	91	85.0	280	JE0217	28k surface antige
3	73	68.2	286	JE0219	28k surface antige
4	61.5	57.5	133	JE0221	28k surface antige
5	60	56.1	278	JE0216	major antigenic pr
6	53	49.5	284	I40882	hypothetical prote
7	47	43.9	160	A12476	hypothetical prote
8	47	43.9	1029	F87369	TonB-dependent rec
9	46	43.0	299	T23932	hypothetical prote
10	46	43.0	1004	JH0470	Na+/K+-exchanging
11	46	43.0	1034	S76134	hypothetical prote
12	45	42.1	160	A75466	2-demethylmenaquin
13	45	42.1	300	S60558	envelope polypepte
14	45	42.1	300	S60526	envelope polypepte
15	45	42.1	490	G85354	hypothetical prote
16	45	42.1	540	S54586	probable membrane
17	44	41.1	1037	A56594	Na+/K+-exchanging
18	44	41.1	1649	C86822	hypothetical prote
19	43.5	40.7	165	G83166	hypothetical prote
20	43	40.2	86	S28126	gas-vesicle protei
21	43	40.2	384	T240502	hypothetical prote
22	43	40.2	422	S73667	adhesin P1 precurs
23	43	40.2	428	S73379	adhesin P1 precurs
24	43	40.2	438	S73952	adhesin P1 precurs
25	43	40.2	459	AH0405	probable cytochrom
26	43	40.2	466	F95307	conserved hypothet
27	43	40.2	547	E81017	ABC transporter, A
28	43	40.2	648	P3BPF6	p3 protein - phage
29	42.5	39.7	534	C82096	aminoacyl-histidin

glutamate synthase	1538	39.7	42.5	30	S73296
probable membrane	277	39.3	42	31	S54501
hypothetical prote	297	39.3	42	32	T21632
UTR2 protein - yea	467	39.3	42	33	S30839
hypothetical prote	631	39.3	42	34	T20037
Na+/K+-exchanging	1038	39.3	42	35	S03632
maltose transport	275	38.8	41.5	36	H90251
probable PRP19-lik	540	38.8	41.5	37	C84744
streptothricine ac	173	38.3	41	38	S65999
ribosomal protein	233	38.3	41	39	T06915
hypothetical prote	264	38.3	41	40	T13743
biotin synthetase	282	38.3	41	41	F64695
hypothetical prote	330	38.3	42	42	H71823
hypothetical prote	331	38.3	41	43	S76408
hypothetical prote	357	38.3	41	44	C72239
hypothetical prote	363	38.3	41	45	T25499
(2'-5')oligo(A) sy	364	38.3	41	46	SYM502
(2'-5')oligo(A) sy	364	38.3	41	47	SYH016
N-acetylglucosaminyl-L	375	38.3	41	48	ACI350
N-acetylglucosaminyl-L	375	38.3	41	49	AF1720
acetyltransferase	377	38.3	41	50	F69008
(2'-5')oligo(A) sy	400	38.3	41	51	SYH018
(2'-5')oligo(A) sy	414	38.3	41	52	SYM503
glutamate/glutamin	426	38.3	41	53	S77606
(2'-5')oligo(A) sy	459	38.3	41	54	A22842
probable cytochrom	466	38.3	41	55	AB0960
probable diacylgly	475	38.3	41	56	H84567
flagellin - Escher	584	38.3	41	57	C48658
outer membrane hem	681	38.3	41	58	E82812
hypothetical prote	756	38.3	41	59	T20109
cell wall surface	893	38.3	41	60	E95053
hypothetical prote	1028	38.3	41	61	T50230
hypothetical prote	1056	38.3	41	62	T00060
probable RTX famil	5188	38.3	41	63	B85547
hypothetical prote	5291	38.3	41	64	F90696
conserved hypothet	267	37.9	40.5	65	C83242
hypothetical prote	855	37.9	40.5	66	T47534
hypothetical prote	118	37.4	40	67	S34346
signal peptidase 2	226	37.4	40	68	A55012
hypothetical prote	233	37.4	40	69	S76857
MT1225-related pro	284	37.4	40	70	H69232
aspartate-semialde	354	37.4	40	71	F64325
cytochrome-c perox	465	37.4	40	72	S47738
cytochrome-c perox	465	37.4	40	73	F91178
cytochrome-c perox	465	37.4	40	74	G86024
xylan 1,4-beta-xy	473	37.4	40	75	JQ1936
aminoacyl-histidin	485	37.4	40	76	H90661
heat shock protein	485	37.4	40	77	F85512
hypothetical prote	543	37.4	40	78	S73595
organic solvent to	792	37.4	40	79	T36966
env polypeptide pr	964	37.4	40	80	B82756
hypothetical prote	964	37.4	40	81	VCLJJC6
hypothetical prote	5627	37.4	40	82	C83339
hypothetical prote	397	36.9	39.5	83	T27950
factor arrest prot	830	36.9	39.5	84	S56940
related to glucan	863	36.9	39.5	85	T49709
hypothetical prote	88	36.4	39	86	A84166
hypothetical prote	117	36.4	39	87	C71886
thymidylate syntha	231	36.4	39	88	T22530
hypothetical prote	286	36.4	39	89	SYBPT4
GTP-binding regula	286	36.4	39	90	G81264
GTP-binding regula	290	36.4	39	91	T02300
pqgB protein - Kle	308	36.4	39	92	S20454
GTP-binding protei	326	36.4	39	93	T16987
GTP-binding regula	326	36.4	39	94	T02340
GTP-binding protei	328	36.4	39	95	T16970
hypothetical prote	348	36.4	39	96	C96530
conserved hypothet	371	36.4	39	97	D69253
p2X4 receptor - hu	388	36.4	39	98	JC6543
cystathionine beta	390	36.4	39	99	AB1656
cystathionine beta	390	36.4	39	100	AG1284

ALIGNMENTS

RESULT 1

28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match 100.0%; Score 107; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKNDGSAIS 20
|||||
DB 59 NTTGVGFLKQKNDGSAIS 78

RESULT 2

28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match 85.0%; Score 91; DB 2; Length 280;
Best Local Similarity 84.2%; Pred. No. 6.1e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKNDGSAIS 19
|||
DB 60 NTTGVGFLKQKNDGSTIS 78

RESULT 3

28k surface antigen 2 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match 68.2%; Score 73; DB 2; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00048;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKNDGSAIS 19
|||||
DB 60 NTTGVGFLKQKNDRCVIS 78

RESULT 4

28k surface antigen 2 - Ehrlichia canis
C:Species: Ehrlichia canis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000
C:Accession: JE0221
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0221
A:Molecule type: DNA
A:Residues: 1-133 <RED>
A:Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC36722.1; PID:g3327966

Query Match 57.5%; Score 61.5; DB 2; Length 133;
Best Local Similarity 68.4%; Pred. No. 0.015;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 TTVGVGFLKQKNDGSAIS 20
|||||
DB 62 TTV-VYGLKENWAGDAISS 79

RESULT 5

28k surface antigen 3 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tri
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0216
A:Molecule type: DNA
A:Residues: 1-278 <RED>
A:Cross-references: GB:AF062761

Query Match 56.1%; Score 60; DB 2; Length 278;
Best Local Similarity 50.0%; Pred. No. 0.056;
Matches 10; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKNDGSAIS 20
|||
DB 60 NPTVALYGLKQDWEGISSSS 79

RESULT 6

major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R:van Vliet, A.H.; Jongejan, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding th
A:Reference number: I40882; MUID:94178956; PMID:8132352
A:Accession: I40882
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C:Genetics:
A:Gene: map1

Query Match 49.5%; Score 53; DB 2; Length 284;

Best Local Similarity 69.2%; Pred. No. 0.76;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGLKQNWGD 15
I :|||:||||
Db 63 TKAVFGLKKWDG 75

RESULT 7

A12476
hypothetical protein all5369 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: A12476
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12476
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077068.1; PID:gl17134508; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all5369

Query Match 43.9%; Score 47; DB 2; Length 160;
Best Local Similarity 88.9%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GVFGKQNW 13
||| ||||
Db 117 GVFSKQNW 125

RESULT 8

F87369
TonB-dependent receptor [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: F87369
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonitskiy, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; MUID:21173698; PMID:11259647
A:Accession: F87369
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1029 <SPO>
A:Cross-references: GB:AE005673; NID:gl3422250; PIDN:AAK22954.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0970

Query Match 43.9%; Score 47; DB 2; Length 1029;
Best Local Similarity 50.0%; Pred. No. 28;
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 NTTGVFGLKQNW--WDGSAIN 20
I : || || | |||| :|
Db 219 NNSGVGLSGNNDWGGSAATN 240

RESULT 9

T23932
hypothetical protein R05D7.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T23932

R:Dobson, R.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19819
A:Accession: T23932
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-299 <WIL>
A:Cross-references: EMBL:Z81105; PIDN:CAB03219.1; GSPDB:GN00019; CESP:R05D7.4
A:Experimental source: clone R05D7
C:Genetics:
A:Gene: CESP:R05D7.4

A:Map position: 1
A:Introns: 29/3; 122/1; 162/1; 231/2; 268/3
C:Superfamily: tropinesterase

Query Match 43.0%; Score 46; DB 2; Length 299;
Best Local Similarity 56.2%; Pred. No. 11;
Matches 9; Conservative 3; Mismatches 2; Indels 2; Gaps 1;

QY 5 GVFGKQNW--GSAI 18
I :||| |||| :|
Db 53 GLFGKQNWNSVGKAL 68

RESULT 10

JH0470
Na+/K+-exchanging ATPase (EC 3.6.3.9) alpha chain (clone pARATNa136) - brine shrimp
C:Species: Artemia franciscana (brine shrimp)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Apr-2002
C:Accession: JH0470; S24196
R:Macias, M.T.; Palmero, I.; Sastre, L.
Gene 105, 197-204, 1991
A:Title: Cloning of a cDNA encoding an Artemia franciscana Na/K ATPase alpha-subunit
A:Reference number: JH0470; MUID:92039032; PMID:1657719
A:Accession: JH0470
A:Molecule type: mRNA
A:Residues: 1-1004 <MAC>
A:Cross-references: EMBL:X56650; NID:gi10933; PIDN:CAA39972.1; PID:gi10934
C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain
C:Keywords: ATP; heterodimer; hydrolase; ion transport; phosphoprotein; potassium transporter
F:2-1004/Product: Na+/K+-transporting ATPase alpha chain #status predicted <NT1>
F:2-75/Domain: intracellular #status predicted <INT1>
F:76-97/Domain: transmembrane #status predicted <TM1>
F:111-130/Domain: transmembrane #status predicted <TM2>
F:131-271/Domain: intracellular #status predicted <INT2>
F:272-296/Domain: transmembrane #status predicted <TM3>
F:301-329/Domain: transmembrane #status predicted <TM4>
F:330-767/Domain: intracellular #status predicted <INT>
F:568-764/Domain: ATPase nucleotide-binding domain homology <ATN>
F:768-791/Domain: transmembrane #status predicted <TM5>
F:830-855/Domain: transmembrane #status predicted <TM6>
F:856-936/Domain: intracellular #status predicted <INT4>
F:937-955/Domain: transmembrane #status predicted <TM7>
F:956-1004/Domain: extracellular #status predicted <EXT>
F:357/Active site: Asp (aspartylphosphate intermediate) #status predicted
F:489/Binding site: Asp (Lys) #status predicted
F:698,702,707/Active site: Asp, Asp, Lys #status predicted

Query Match 43.0%; Score 46; DB 2; Length 1004;
Best Local Similarity 40.0%; Pred. No. 39;
Matches 6; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 6 VFGKQNWDSAIN 20
:|||:| | :|
Db 864 LFGLKKHWSRAVND 878

RESULT 11

S76134
hypothetical protein - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S76134
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76134

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1034 <KAN>

A:Cross-references: EMBL:D90914; GB:AB001339; NID:g1653477; PIDN:BAAL8393.1; PID:g165348
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: phosphoenolpyruvate carboxylase

Query Match 43.0%; Score 46; DB 2; Length 1034;

Best Local Similarity 50.0%; Pred. No. 41;

Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 NTTVGVFLKQKQNDGS 16

Db 2 NLAVPAPGLSTNWSGN 17

RESULT 12

A:2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) DR0859 [similarity] - Deinococcus

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000

C:Accession: A75466

R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: A75466

A:Molecule type: DNA

A:Residues: 1-160 <WHI>

A:Cross-references: GB:AE001940; GB:AEO05113; NID:g6458577; PIDN:AAF10437.1; PID:g645857

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0859

A:Map position: 1

C:Keywords: methyltransferase

Query Match 42.1%; Score 45; DB 2; Length 160;

Best Local Similarity 46.7%; Pred. No. 7.8;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 4 VGVFLKQKQNDGSAI 18

Db 79 LGVFGVNGWEGVII 93

RESULT 13

A:envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-330-1 and o

C:Species: human immunodeficiency virus type 1, HIV-1

A:Variety: isolate CI-330-1; isolate CI-330-2; isolate CI-330-4

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999

C:Accession: S60558; S60559; S60560

R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J.;

AIDS 8, 21-26, 1994

A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cote d

A:Reference number: S60521; MUID:94280700; PMID:8011235

A:Accession: S60558

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-300 <JAN>

A:Cross-references: EMBL:X72062; NID:g468697; PIDN:CAA50943.1; PID:g468698

A:Experimental source: isolate CI-330-1

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993
 A:Accession: S60559

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-300 <JAN>

A:Cross-references: EMBL:X72063; NID:g468699; PIDN:CAA50944.1; PID:g468700

A:Experimental source: isolate CI-330-2

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

A:Accession: S60560

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-300 <JAN>

A:Cross-references: EMBL:X72065; NID:g468701; PIDN:CAA50946.1; PID:g468702

A:Experimental source: isolate CI-330-4

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 42.1%; Score 45; DB 2; Length 300;

Best Local Similarity 47.1%; Pred. No. 15;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TVGVFLKQKQNDGSAIS 19

Db 117 TSGLFNSTWNNNGTAIT 133

RESULT 14

A:envelope polyprotein gp41 - human immunodeficiency virus type 1 (isolate CI-330-3) (f

C:Species: human immunodeficiency virus type 1, HIV-1

A:Variety: isolate CI-330-3

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 26-Aug-1999

C:Accession: S60536

R:Janssens, W.; Heyndrickx, L.; Van de Peer, Y.; Bouckaert, A.; Fransen, K.; Motte, J

AIDS 8, 21-26, 1994

A:Title: Molecular phylogeny of part of the env gene of HIV-1 strains isolated in Cct

A:Reference number: S60536; MUID:94280700; PMID:8011235

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-300 <JAN>

A:Cross-references: EMBL:X72064; NID:g468784; PIDN:CAA50945.1; PID:g468785

A:Experimental source: isolate CI-330-3

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C:Genetics:

A:Gene: env

C:Superfamily: type E retrovirus env polyprotein

C:Keywords: AIDS; glycoprotein; immunodeficiency; polyprotein

Query Match 42.1%; Score 45; DB 2; Length 300;

Best Local Similarity 47.1%; Pred. No. 15;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 TVGVFLKQKQNDGSAIS 19

Db 117 TSGLFNSTWNNNGTAIT 133

RESULT 15

A:hypothetical protein AT4930340 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: G85354

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: G85354

A>Status: preliminary

Db 93 TLMQVWGLRVQNRDGSALS 112

RESULT 20

S28126
gas-vesicle protein gvpM - Haloferax mediterranei
C:Species: Haloferax mediterranei
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C:Accession: S28126
R:Englert, C.; Krueger, K.; Offner, S.; Pfeifer, F.
J. Mol. Biol. 227, 586-592, 1992
A:Title: Three different but related gene clusters encoding gas vesicles in halophilic
A:Reference number: S28113; MUID:93021102; PMID:1404376
A:Accession: S28126
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-86 <ENG>
A:CROSS-references: EMBL:X64701; NID:g58346; PIDN:CAA45955.1; PID:g58360
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
C:Genetics:
A:Gene: gvpM

Query Match 40.2%; Score 43; DB 2; Length 86;
Best Local Similarity 40.0%; Pred. No. 8.3;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTVGVFGLKQNDGS 16
||: ||: |||:
Db 51 TMTAYGMFENNDAT 65

RESULT 21

T40502
hypothetical protein SPBC4F6.05c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40502
R:Gwilliam, R.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z21933
A:Accession: T40502
A:Status: preliminary; translated from GB/EMBL/DDB/J
A:Molecule type: DNA
A:Residues: 1-384 <GWI>
A:CROSS-references: EMBL:AL031534; PIDN:CAA20725.1; GSPDB:GN00067; SPDB:SPBC4F6.05c
A:Experimental source: strain 972h; cosmid c4F6
C:Genetics:
A:Gene: SPDB:SPBC4F6.05c
A:Map position: 2

Query Match 40.2%; Score 43; DB 2; Length 384;
Best Local Similarity 57.1%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 VFGLKQNDGSALS 19
||| ||| ||
Db 110 VFGASDKWDGLLIS 123

RESULT 22

S73667
adhesin P1 precursor homolog P02_orf422V - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73667
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73667
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-422 <HIM>
A:CROSS-references: EMBL:AE000032; GB:U00089; NID:gl674011; PIDN:AAB95988.1; PID:g167
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
A:Start codon: GTG

Query Match 40.2%; Score 43; DB 2; Length 422;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFGLK 10
|||: ||: |||:
Db 141 NTTIGAYGLK 150

RESULT 23

S73379
adhesin P1 precursor homolog C09_orf428V - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73379
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73379
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-428 <HIM>
A:CROSS-references: EMBL:AE000006; GB:U00089; NID:gl673695; PIDN:AAB95701.1; PID:g167
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
A:Start codon: GTG

Query Match 40.2%; Score 43; DB 2; Length 428;
Best Local Similarity 70.0%; Pred. No. 47;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFGLK 10
|||: ||: |||:
Db 141 NTTIGAYGLK 150

RESULT 24

S73952
adhesin P1 precursor homolog GT9_orf438V - Mycoplasma pneumoniae (strain ATCC 29342)
N:Alternate names: hypothetical protein GT9_orf438V
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73952
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73952
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-438 <HIM>
A:CROSS-references: EMBL:AE000060; GB:U00089; NID:gl674327; PIDN:AAB96274.1; PID:g167
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
A:Start codon: GTG

Query Match 40.2%; Score 43; DB 2; Length 438;
Best Local Similarity 70.0%; Pred. No. 49;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFGLK 10

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: S82096
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
 Chardson, D.; Encheva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, J.
 L. R.R.; Mekalanav, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406: 477-483, 2001

A:Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
 A:Reference number: A82035; MUID:20406833; PMID:10952301
 A:Accession: S82096
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1534
 A:Cross-references: GB:AE004299; GB:AE003852; NID:g9656835; PIDN:AAF95423.1; GSPDB:GN001
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor
 C:Genetics:
 A:Gene: VC2279
 A:Map position: 1

Query Match 39.7%; Score 42.5; DB 2; Length 534;
 Best Local Similarity 47.1%; Pred. No. 72;
 Matches 8; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 5 GVFGKONW-DGSAISN 20
 Db 199 GAFGLKQWLEGDILLN 215

RESULT 30

S73296
 glutamate synthase (ferredoxin) (EC 1.4.7.1) precursor - red alga (*Porphyra purpurea*) ch
 N:Alternate names: GOGAT enzyme gltB
 C:Species: chloroplast *Porphyra purpurea*
 C:Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 04-Mar-2000
 C:Accession: S73296
 R:Reith, M.; Munholland, J.
 Plant Mol. Biol. Rep. 13: 333-335, 1995
 A:Title: Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome.
 A:Reference number: S73108
 A:Accession: S73296
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1538 <RRI>
 A:Cross-references: EMBL:U38804; NID:gl276652; PIDN:AA08261.1; PID:gl276841
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
 C:Genetics:
 A:Gene: gltB
 A:Genome: chloroplast
 C:Superfamily: glutamate synthase (NADPH)
 C:Keywords: 3Fe-4S; chloroplast; metalloprotein; oxidoreductase
 F:1-33/Domain: propeptide #status predicted <PRO>
 F:34-1538/Product: glutamate synthase #status predicted <MAT>
 F:34/Active site: Cys #status predicted
 F:1162,1168,1173/Binding site: 3Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 39.7%; Score 42.5; DB 2; Length 1538;
 Best Local Similarity 47.1%; Pred. No. 2.3e+02;
 Matches 9; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY 1 NTVGVF-----GLKQWDSAI 18
 Db 345 NTEISDFEYVYSLQEPWDGPA 367

RESULT 31

S54501
 Probable membrane protein YPR027c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein YP9367.07c
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 19-Apr-2002
 C:Accession: S54501
 R:Badcock, K.; Churcher, C.M.
 submitted to the EMBL Data Library, May 1995
 A:Reference number: S54059

A:Accession: S54501
 A:Molecule type: DNA
 A:Residues: 1-277 <BAD>
 A:Cross-references: EMBL:Z49274; NID:g809585; PID:g809592; GSPDB:GN00016; MIPS:YPR027
 A:Experimental source: strain AB9722h
 C:Genetics:
 A:Gene: MIPS:YPR027c
 A:Cross-references: SGD:S0006231
 A:Map position: 16R
 C:Keywords: transmembrane protein
 F:4-20/Domain: transmembrane #status predicted <TM1>
 F:86-102/Domain: transmembrane #status predicted <TM2>

Query Match 39.3%; Score 42; DB 2; Length 277;
 Best Local Similarity 75.0%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 11 QNWDGSAI 18
 Db 246 ENWDGSAV 253

RESULT 32

T21632
 hypothetical protein F32B4.6 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
 C:Accession: T21632
 R:White, S.
 submitted to the EMBL Data Library, November 1996

A:Reference number: Z19452
 A:Accession: T21632
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-297 <WIL>
 A:Cross-references: EMBL:Z81522; PIDN:CA804232.1; GSPDB:GN00019; CESP:F32B4.6
 A:Experimental source: clone F32B4
 C:Genetics:
 A:Gene: CESP:F32B4.6
 A:Map position: 1
 A:Introns: 20/3; 152/1; 256/3
 C:Superfamily: tropinesterase

Query Match 39.3%; Score 42; DB 2; Length 297;
 Best Local Similarity 66.7%; Pred. No. 46;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 GVFGKONW 13
 Db 43 GLFGTKENW 51

RESULT 33

S30839
 UTR2 protein - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: protein YEL040w
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 06-Feb-1998
 C:Accession: S30839; S50504; S38545
 R:Mulligan, J.T.; Dietrich, F.S.; Hennessey, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylo
 submitted to the EMBL Data Library, February 1993
 A:Reference number: S30812
 A:Accession: S30839
 A:Molecule type: DNA
 A:Residues: 1-467 <MUL>
 A:Cross-references: GB:U18779; EMBL:L10830; NID:g603625; PID:g603639
 R:Dietrich, F.S.
 submitted to the EMBL Data Library, December 1994

A:Description: The sequence of *S. cerevisiae* cosmids 8199, 8334, and 9871.
 A:Reference number: S50491
 A:Accession: S50504
 A:Molecule type: DNA
 A:Residues: 1-467 <DIE>

A;Accession: S07049
A;Molecule type: mRNA
A;Residues: 397-521 <VAR>
A;Cross-references: EMBL:X17471
A;Note: the authors translated the codon ACC for residue 3 as Asn and AAT for residue 4.
C;Genetics:
A;Gene: FlyBase:Atp-alpha
A;Cross-references: FlyBase:FBgn0002921
A;Map position: 3R 93B
C;Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain
C;Keywords: ATP; heterodimer; hydrolase; ion transport; phosphoprotein; potassium transporter
F;113-135/Domain: transmembrane #status predicted <TM1>
F;146-165/Domain: transmembrane #status predicted <TM2>
F;166-305/Domain: intracellular #status predicted <INT2>
F;306-328/Domain: transmembrane #status predicted <TM3>
F;335-363/Domain: transmembrane #status predicted <TM4>
F;364-801/Domain: intracellular #status predicted <INT3>
F;602-798/Domain: ATPase nucleotide-binding domain homology <ATN>
F;802-825/Domain: transmembrane #status predicted <TM5>
F;864-889/Domain: transmembrane #status predicted <TM6>
F;890-966/Domain: intracellular #status predicted <INT4>
F;967-993/Domain: transmembrane #status predicted <TM7>
F;994-1038/Domain: extracellular #status predicted <EXT>
F;391/Active site: Asp (aspartylphosphate intermediate) #status predicted
F;523/Binding site: ATP (Lys) #status predicted
F;732,736,741/Active site: Asp, Asp, Lys #status predicted

Query Match 39.3%; Score 42; DB 1; Length 1038;
Best Local Similarity 33.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 6 VFGLKQNWGSAISN 20
:|::: |||:::
Db 898 LFGIRKMWDSKAVND 912

Search completed: October 6, 2003, 07:49:30
Job time : 16.7463 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 11.6418 Seconds
(without alignments)
72.688 Million cell updates/sec

Title: US-09-765-739A-3

Perfect score: 107

Sequence: 1 NTTVGVFLKQNWGSAISN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued_Patents_AA:*

1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep:*

2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep:*

3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep:*

4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep:*

5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep:*

6: /cgn2.6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	276	3	US-08-953-326-18
2	107	100.0	280	3	US-08-733-230-4
3	107	100.0	280	3	US-08-953-326-4
4	107	100.0	281	4	US-09-660-587-9
5	107	100.0	281	4	US-09-261-358A-9
6	107	100.0	281	4	US-09-201-458-5
7	107	100.0	281	4	US-09-314-701-2
8	91	85.0	280	3	US-08-953-326-17
9	88	82.2	280	4	US-09-660-587-14
10	88	82.2	280	4	US-09-261-358A-14
11	88	82.2	280	4	US-09-201-458-10
12	88	82.2	280	4	US-09-314-701-12
13	82	76.6	288	4	US-09-314-701-32
14	80	74.8	280	4	US-09-660-587-6
15	80	74.8	280	4	US-09-261-358A-6
16	80	74.8	280	4	US-09-314-701-38
17	73	68.2	286	3	US-08-953-326-15
18	73	68.2	286	4	US-09-660-587-12
19	73	68.2	286	4	US-09-261-358A-12
20	73	68.2	286	4	US-09-201-458-8
21	73	68.2	286	4	US-09-314-701-8
22	72	67.3	278	4	US-09-660-587-2
23	72	67.3	278	4	US-09-261-358A-2
24	72	67.3	278	4	US-09-201-458-2
25	72	67.3	307	4	US-09-314-701-36
26	68	63.6	276	4	US-09-660-587-44
27	68	63.6	276	4	US-09-314-701-42

28	61.5	57.5	133	3	US-08-953-326-20	Sequence 20, Appl
29	61.5	57.5	133	4	US-09-660-587-7	Sequence 7, Appl
30	61.5	57.5	133	4	US-09-261-358A-7	Sequence 7, Appl
31	61.5	57.5	133	4	US-09-201-458-3	Sequence 3, Appl
32	61.5	57.5	283	4	US-09-660-587-4	Sequence 4, Appl
33	61.5	57.5	283	4	US-09-261-358A-4	Sequence 4, Appl
34	60	56.1	278	3	US-08-953-326-16	Sequence 16, Appl
35	60	56.1	278	4	US-09-660-587-13	Sequence 13, Appl
36	60	56.1	278	4	US-09-261-358A-13	Sequence 13, Appl
37	60	56.1	278	4	US-09-201-458-9	Sequence 9, Appl
38	60	56.1	278	4	US-09-314-701-10	Sequence 10, Appl
39	60	56.1	280	4	US-09-660-587-11	Sequence 11, Appl
40	60	56.1	280	4	US-09-261-358A-11	Sequence 11, Appl
41	60	56.1	280	4	US-09-201-458-7	Sequence 7, Appl
42	60	56.1	280	4	US-09-314-701-6	Sequence 6, Appl
43	53	49.5	284	4	US-09-660-587-15	Sequence 15, Appl
44	53	49.5	284	4	US-09-261-358A-15	Sequence 15, Appl
45	53	49.5	284	4	US-09-201-458-11	Sequence 11, Appl
46	51	47.7	287	3	US-08-733-230-2	Sequence 2, Appl
47	51	47.7	287	3	US-08-953-326-2	Sequence 2, Appl
48	48	44.9	308	4	US-09-584-568C-8	Sequence 8, Appl
49	46	43.0	299	4	US-09-584-568C-6	Sequence 6, Appl
50	43.5	40.7	202	4	US-09-252-991A-22713	Sequence 22713, A
51	43	40.2	956	4	US-09-134-078-63	Sequence 63, Appl
52	42	39.3	297	4	US-09-584-568C-7	Sequence 7, Appl
53	42	39.3	447	4	US-09-252-991A-25916	Sequence 25916, A
54	41	38.3	365	3	US-08-840-146-19	Sequence 19, Appl
55	41	38.3	365	3	US-09-360-220-19	Sequence 19, Appl
56	41	38.3	400	2	US-08-436-771-11	Sequence 11, Appl
57	41	38.3	400	2	US-08-434-998-11	Sequence 11, Appl
58	41	38.3	400	5	PCT-US95-02058-11	Sequence 11, Appl
59	41	38.3	400	5	US-09-946-678-2	Sequence 2, Appl
60	41	38.3	684	4	US-09-252-991A-21630	Sequence 21630, A
61	40.5	37.9	268	4	US-09-252-991A-23574	Sequence 23574, A
62	40	37.4	96	4	US-09-252-991A-26356	Sequence 26356, A
63	40	37.4	430	4	US-09-252-991A-23547	Sequence 23547, A
64	40	37.4	2284	4	US-08-630-822A-64	Sequence 64, Appl
65	39	36.4	137	2	US-09-005-069-64	Sequence 64, Appl
66	39	36.4	137	2	US-09-171-156A-23	Sequence 23, Appl
67	39	36.4	137	4	US-09-004-730A-23	Sequence 23, Appl
68	39	36.4	137	4	US-08-981-799A-23	Sequence 23, Appl
69	39	36.4	203	4	US-09-134-001C-5565	Sequence 5565, Ap
70	39	36.4	291	4	US-09-443-184-54	Sequence 54, Appl
71	39	36.4	301	4	US-09-328-352-7439	Sequence 7439, Ap
72	39	36.4	388	2	US-08-742-621-1	Sequence 1, Appl
73	39	36.4	388	3	US-09-191-608-22	Sequence 22, Appl
74	39	36.4	461	4	US-09-134-001C-5311	Sequence 5311, Ap
75	39	36.4	468	4	US-09-252-991A-18608	Sequence 18608, A
76	39	36.4	522	4	US-09-252-991A-16994	Sequence 16994, A
77	39	36.4	527	1	US-08-462-484-10	Sequence 10, Appl
78	39	36.4	527	1	US-08-441-147-10	Sequence 10, Appl
79	39	36.4	527	5	PCT-US95-07536-10	Sequence 10, Appl
80	39	36.4	2476	2	US-08-276-967-2	Sequence 2, Appl
81	39	36.4	283	4	US-09-314-701-40	Sequence 40, Appl
82	38.5	36.0	185	4	US-09-134-001C-3293	Sequence 3293, Ap
83	38	35.5	373	4	US-09-071-035-116	Sequence 116, Appl
84	38	35.5	377	1	US-07-772-087-2	Sequence 2, Appl
85	38	35.5	406	4	US-09-071-035-114	Sequence 114, App
86	38	35.5	697	4	US-09-328-352-6212	Sequence 6212, Ap
87	38	35.5	738	3	US-08-989-385-1	Sequence 1, Appl
88	38	35.5	738	3	US-09-593-826-1	Sequence 1, Appl
89	38	35.5	137	4	US-09-732-210-638	Sequence 638, App
90	37.5	35.0	137	1	US-08-190-802A-70	Sequence 70, Appl
91	37	34.6	31	1	US-08-190-802A-129	Sequence 129, App
92	37	34.6	31	1	US-08-190-802A-173	Sequence 173, App
93	37	34.6	31	3	US-08-477-346-70	Sequence 70, Appl
94	37	34.6	31	3	US-08-477-346-129	Sequence 129, App
95	37	34.6	31	3	US-08-477-346-173	Sequence 173, App
96	37	34.6	31	4	US-08-473-089-70	Sequence 70, Appl
97	37	34.6	31	4	US-08-473-089-129	Sequence 129, App
98	37	34.6	31	4	US-08-473-089-173	Sequence 173, App
99	37	34.6	31	4	US-08-487-072A-70	Sequence 70, Appl
100	37	34.6	31	4		

ALIGNMENTS

```
RESULT 1
US-08-953-326-18
; Sequence 18, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match 100.0%; Score 107; DB 3; Length 276;
Best Local Similarity 100.0%; Pred. No. 7.7e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNWGSAISN 20
Db 59 NTTGVFGLKQNWGSAISN 78

RESULT 2
US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match 100.0%; Score 107; DB 3; Length 280;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNWGSAISN 20
Db 60 NTTGVFGLKQNWGSAISN 79

RESULT 3
US-08-953-326-4
; Sequence 4, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match 100.0%; Score 107; DB 3; Length 280;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNWGSAISN 20
Db 60 NTTGVFGLKQNWGSAISN 79

RESULT 4
US-09-660-587-9
; Sequence 9, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/733,230
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-660-587-9

Query Match 100.0%; Score 107; DB 3; Length 280;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNWGSAISN 20
Db 60 NTTGVFGLKQNWGSAISN 79
```

```
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-4
```

```
Query Match 100.0%; Score 107; DB 3; Length 280;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNWGSAISN 20
Db 60 NTTGVFGLKQNWGSAISN 79
```

```
RESULT 3
US-08-953-326-4
; Sequence 4, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4
```

```
Query Match 100.0%; Score 107; DB 3; Length 280;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNWGSAISN 20
Db 60 NTTGVFGLKQNWGSAISN 79
```

```
RESULT 4
US-09-660-587-9
; Sequence 9, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/733,230
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-660-587-9
```

; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-660-587-9

Query Match 100.0%; Score 107; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 7.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|||||
DB 59 NTTGVFGLKQNWGSAISN 78

RESULT 5
US-09-261-358A-9
; Sequence 9, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-261-358A-9

Query Match 100.0%; Score 107; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 7.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|||||
DB 59 NTTGVFGLKQNWGSAISN 78

RESULT 6
US-09-201-458-5
; Sequence 5, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 281
; TYPE: PRT

; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-201-458-5

Query Match 100.0%; Score 107; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 7.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|||||
DB 59 NTTGVFGLKQNWGSAISN 78

RESULT 7
US-09-314-701-2
; Sequence 2, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-2

Query Match 100.0%; Score 107; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 7.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|||||
DB 59 NTTGVFGLKQNWGSAISN 78

RESULT 8
US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match 85.0%; Score 91; DB 3; Length 280;

Best Local Similarity 84.2%; Pred. No. 3.3e-07;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSGTIS 78

RESULT 9

US-09-660-587-14
; Sequence 14, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-660-587-14

Query Match 82.2%; Score 88; DB 4; Length 280;
Best Local Similarity 84.2%; Pred. No. 1e-06; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSGTIS 78

RESULT 10

US-09-261-358A-14
; Sequence 14, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-261-358A-14

Query Match 82.2%; Score 88; DB 4; Length 280;
Best Local Similarity 84.2%; Pred. No. 1e-06; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSGTIS 78

RESULT 11

US-09-201-458-10
; Sequence 10, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 10
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-201-458-10

Query Match 82.2%; Score 88; DB 4; Length 280;
Best Local Similarity 84.2%; Pred. No. 1e-06; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSGTIS 78

RESULT 12

US-09-314-701-12
; Sequence 12, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-12

Query Match 82.2%; Score 88; DB 4; Length 280;
Best Local Similarity 84.2%; Pred. No. 1e-06; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
Db 60 NTTGVFGLKQNDGSGTIS 78

RESULT 13

US-09-314-701-32
; Sequence 32, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19

; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-32

Query Match 76.6%; Score 82; DB 4; Length 288;
Best Local Similarity 70.0%; Pred. No. 1e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|:|||||:|:|
DB 60 NTTGVFGLKQWDGATIKD 79

RESULT 14
US-09-660-587-6

; Sequence 6, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-6 protein
US-09-660-587-6

Query Match 74.8%; Score 80; DB 4; Length 280;
Best Local Similarity 70.0%; Pred. No. 2.1e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|:|||||:|:|
DB 59 NSTGVFGLKHDWNGGTISN 78

RESULT 15
US-09-261-358A-6

; Sequence 6, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28SA3 protein
US-09-261-358A-6

Query Match 74.8%; Score 80; DB 4; Length 280;
Best Local Similarity 70.0%; Pred. No. 2.1e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|:|||||:|:|
DB 59 NSTGVFGLKHDWNGGTISN 78

RESULT 16

US-09-314-701-38
; Sequence 38, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-38

Query Match 74.8%; Score 80; DB 4; Length 280;
Best Local Similarity 70.0%; Pred. No. 2.1e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
|:|||||:|:|
DB 59 NSTGVFGLKHDWNGGTISN 78

RESULT 17

US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 68.2%; Score 73; DB 3; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0003;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19

```
Db      ~|||||||:|:|:|  ||
        60 NTTGVFGIEQDWDRCVIS 78

RESULT 18
US-09-660-587-12
; Sequence 12, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-660-587-12

Query Match      68.2%; Score 73; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0003;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 NTTGVFGGLKQNDGSAIS 19
        |||||||:|:|:|  ||
Db      60 NTTGVFGIEQDWDRCVIS 78

RESULT 19
US-09-261-358A-12
; Sequence 12, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-261-358A-12

Query Match      68.2%; Score 73; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0003;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 NTTGVFGGLKQNDGSAIS 19
        |||||||:|:|:|  ||
Db      60 NTTGVFGIEQDWDRCVIS 78

RESULT 20
US-09-201-458-8
; Sequence 8, Application US/09201458A
```

```
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-201-458-8

Query Match      68.2%; Score 73; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0003;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 NTTGVFGGLKQNDGSAIS 19
        |||||||:|:|:|  ||
Db      60 NTTGVFGIEQDWDRCVIS 78

RESULT 21
US-09-314-701-8
; Sequence 8, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517to
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-314-701-8

Query Match      68.2%; Score 73; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0003;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 NTTGVFGGLKQNDGSAIS 19
        |||||||:|:|:|  ||
Db      60 NTTGVFGIEQDWDRCVIS 78

RESULT 22
US-09-660-587-2
; Sequence 2, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
```

```
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein
US-09-660-587-2

Query Match          67.3%; Score 72; DB 4; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00042;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTGVFGLKQNDGSAI 18
Db 61 STGVFGLKHDWDGSP1 77

RESULT 23
US-09-261-358A-2
; Sequence 2, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28-1 protein
US-09-261-358A-2

Query Match          67.3%; Score 72; DB 4; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00042;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTGVFGLKQNDGSAI 18
Db 61 STGVFGLKHDWDGSP1 77

RESULT 24
US-09-201-458-2
; Sequence 2, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a 30 kDa immunoreactive
; OTHER INFORMATION: protein of Ehrlichia canis
US-09-201-458-2
```

```
Query Match          67.3%; Score 72; DB 4; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00042;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTGVFGLKQNDGSAI 18
Db 61 STGVFGLKHDWDGSP1 77

RESULT 25
US-09-314-701-36
; Sequence 36, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; OTHER INFORMATION:
US-09-314-701-36

Query Match          67.3%; Score 72; DB 4; Length 307;
Best Local Similarity 76.5%; Pred. No. 0.00048;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTGVFGLKQNDGSAI 18
Db 90 STGVFGLKHDWDGSP1 106

RESULT 26
US-09-660-587-44
; Sequence 44, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 44
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-3 protein
US-09-660-587-44

Query Match          63.6%; Score 68; DB 4; Length 276;
Best Local Similarity 61.1%; Pred. No. 0.0019;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTVGVFGLKQNDGSAI 18
Db 60 NTTTGIFGLKESWTGII 77

RESULT 27
US-09-314-701-42
```

```
; Sequence 42, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 634451710
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; OTHER INFORMATION: partial amino acid sequence of p28-5 protein
US-09-314-701-42

Query Match      63.6%; Score 68; DB 4; Length 276;
Best Local Similarity 61.1%; Pred. No. 0.0019; 4; Indels 0; Gaps 0;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQNDGSAI 18
   ||| |::||::||::||
Db 60 NTTGIFGLKESWTGGII 77

RESULT 28
US-08-953-326-20
; Sequence 20, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Susan M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: US/167CI
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; OTHER INFORMATION: partial amino acid sequence of ECa28SA2 protein
US-08-953-326-20

Query Match      57.5%; Score 61.5; DB 3; Length 133;
Best Local Similarity 58.4%; Pred. No. 0.0095; 3; Mismatches 1; Gaps 1;
Matches 13; Conservative

Qy 2 TTGVFGLKQNDGSAISN 20
   ||| |::||::||::||
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 29
US-09-660-587-7
; Sequence 7, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 3
```

```
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 7
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: partial amino acid sequence of p28-5 protein
US-09-660-587-7

Query Match      57.5%; Score 61.5; DB 4; Length 133;
Best Local Similarity 68.4%; Pred. No. 0.0095; 3; Mismatches 1; Gaps 1;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 2 TTGVFGLKQNDGSAISN 20
   ||| |::||::||::||
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 30
US-09-261-358A-7
; Sequence 7, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 7
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: partial amino acid sequence of ECa28SA2 protein
US-09-261-358A-7

Query Match      57.5%; Score 61.5; DB 4; Length 133;
Best Local Similarity 68.4%; Pred. No. 0.0095; 3; Mismatches 1; Gaps 1;
Matches 13; Conservative

Qy 2 TTGVFGLKQNDGSAISN 20
   ||| |::||::||::||
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 31
US-09-201-458-3
; Sequence 3, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 3
```

```

; LENGTH: 133
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis 28-kDa protein-1
; US-09-201-458-3

Query Match          57.5%; Score 61.5; DB 4; Length 133;
Best Local Similarity 68.4%; Pred. No. 0.0095; 2; Indels 1; Gaps 1;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 TTVGVFGLKONWDGSAISN 20
   ||| |::||::|| |::|
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 32
US-09-660-587-4
; Sequence 4, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; EARLIER FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 4
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-5 protein
; US-09-660-587-4

Query Match          57.5%; Score 61.5; DB 4; Length 283;
Best Local Similarity 68.4%; Pred. No. 0.023; 2; Indels 1; Gaps 1;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 TTVGVFGLKONWDGSAISN 20
   ||| |::||::|| |::|
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 33
US-09-261-358A-4
; Sequence 4, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 4
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28SA2 protein
; US-09-261-358A-4

```

```

Query Match          57.5%; Score 61.5; DB 4; Length 283;
Best Local Similarity 68.4%; Pred. No. 0.023; 2; Indels 1; Gaps 1;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 TTVGVFGLKONWDGSAISN 20
   ||| |::||::|| |::|
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 34
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: BurrIDGE, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; US-08-953-326-16

Query Match          56.1%; Score 60; DB 3; Length 278;
Best Local Similarity 50.0%; Pred. No. 0.039; 4; Indels 0; Gaps 0;
Matches 10; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKONWDGSAISN 20
   ||| |::||::|| |::|
Db 60 NPTVALYGLKQDWEGISSSS 79

RESULT 35
US-09-660-587-13
; Sequence 13, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
; US-09-660-587-13

Query Match          56.1%; Score 60; DB 4; Length 278;

```

Mon Oct 6 09:43:03 2003

us-09-765-739a-3.ra1

Page 10

Best Local Similarity 50.0%; Pred. No. 0.039;
Matches 10; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTVGVGGLKQWDCSAISN 20
Db 60 NPTVALYGLKQWEGISSSS 79

Search completed: October 6, 2003, 07:53:51
Job time : 13.6418 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 25.806 Seconds
(without alignments)
189.995 Million cell updates/sec

Title: US-09-765-739A-5

Perfect score: 105

Sequence: 1 NTTGVFGIEQDWDRCVTS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organalle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	286	2	052105 ehrlichia c
2	105	100.0	287	2	Q8G9U3 ehrlichia c
3	105	100.0	291	2	Q8G9U2 ehrlichia c
4	105	100.0	291	2	Q8G921 ehrlichia c
5	100	95.2	291	2	Q8G8P3 ehrlichia c
6	74	70.5	280	2	052107 ehrlichia c
7	73	69.5	281	2	Q93DD2 ehrlichia c
8	73	69.5	281	2	Q9ACI9 ehrlichia c
9	71	67.6	288	2	Q9ZGJ2 ehrlichia c
10	70	66.7	246	2	Q9RH35 ehrlichia c
11	70	66.7	275	2	Q93DD4 ehrlichia c
12	70	66.7	276	2	Q93DD1 ehrlichia c
13	70	66.7	276	2	Q85817 ehrlichia c
14	70	66.7	276	2	Q8G9U0 ehrlichia c
15	69	65.7	280	2	Q93DD3 ehrlichia c
16	69	65.7	280	2	Q9ZGM9 ehrlichia c

17	69	65.7	280	2	085816	085816 ehrlichia c
18	62	59.0	280	2	Q9ADV3	Q9ADV3 ehrlichia c
19	62	59.0	280	2	Q9F473	Q9F473 ehrlichia c
20	58	55.2	278	2	Q9F472	Q9F472 ehrlichia c
21	58	55.2	278	2	Q9R8A7	Q9R8A7 ehrlichia c
22	58	55.2	278	2	Q9R8A7	Q9R8A7 ehrlichia c
23	58	55.2	278	2	Q9R3J3	Q9R3J3 ehrlichia c
24	58	55.2	278	2	Q9R8A6	Q9R8A6 ehrlichia c
25	58	55.2	278	2	Q9R8A9	Q9R8A9 ehrlichia c
26	58	55.2	278	2	Q9R8A5	Q9R8A5 ehrlichia c
27	58	55.2	307	2	Q9ZGJ1	Q9ZGJ1 ehrlichia c
28	54	51.4	276	2	Q9F475	Q9F475 ehrlichia c
29	51	48.6	276	2	Q8G948	Q8G948 ehrlichia c
30	51	48.6	277	2	Q8GGU1	Q8GGU1 ehrlichia c
31	51	48.6	277	2	Q8G8W7	Q8G8W7 ehrlichia c
32	51	48.6	278	2	052106	052106 ehrlichia c
33	51	48.6	278	2	Q8G8Q5	Q8G8Q5 ehrlichia c
34	50	47.6	280	2	052104	052104 ehrlichia c
35	50	47.6	280	2	Q8GGU5	Q8GGU5 ehrlichia c
36	50	47.6	280	2	Q8G8J3	Q8G8J3 ehrlichia c
37	46	43.8	272	2	Q9AMF6	Q9AMF6 ehrlichia s
38	46	43.8	272	2	Q93E54	Q93E54 cowdria rum
39	46	43.8	280	2	Q93E55	Q93E55 cowdria rum
40	46	43.8	280	2	Q93E58	Q93E58 cowdria rum
41	46	43.8	284	2	Q9AFAL	Q9AFAL cowdria rum
42	46	43.8	284	2	Q46327	Q46327 cowdria rum
43	45	42.9	253	16	Q93017	Q93017 rhizobium m
44	45	42.9	290	2	Q9AEU3	Q9AEU3 cowdria rum
45	44.5	42.4	91	2	Q9L5F8	Q9L5F8 salmonella
46	44.5	42.4	92	16	Q935M7	Q935M7 salmonella
47	44	41.9	135	5	Q95275	Q95275 osteretia
48	44	41.9	172	16	Q9CKK3	Q9CKK3 pasteurilla
49	44	41.9	253	16	Q8KAQ8	Q8KAQ8 chlorobium
50	44	41.9	265	2	Q9AF99	Q9AF99 cowdria rum
51	44	41.9	270	2	Q9AF98	Q9AF98 cowdria rum
52	44	41.9	275	2	Q93E59	Q93E59 cowdria rum
53	44	41.9	276	2	Q93E60	Q93E60 cowdria rum
54	44	41.9	276	2	Q93E53	Q93E53 cowdria rum
55	44	41.9	277	2	Q93E55	Q93E55 cowdria rum
56	44	41.9	278	2	Q93E52	Q93E52 cowdria rum
57	44	41.9	278	2	Q93E57	Q93E57 cowdria rum
58	44	41.9	278	2	Q93E56	Q93E56 cowdria rum
59	44	41.9	287	2	Q46329	Q46329 cowdria rum
60	44	41.9	287	2	Q9R425	Q9R425 cowdria rum
61	44	41.9	287	2	Q46331	Q46331 cowdria rum
62	44	41.9	290	2	Q46324	Q46324 cowdria rum
63	44	41.9	290	2	Q93E64	Q93E64 cowdria rum
64	44	41.9	290	2	Q46333	Q46333 cowdria rum
65	44	41.9	290	2	Q46332	Q46332 cowdria rum
66	44	41.9	290	2	Q46330	Q46330 cowdria rum
67	44	41.9	305	11	Q8CHM9	Q8CHM9 rattus norv
68	44	41.9	702	11	Q8C3X8	Q8C3X8 mus musculu
69	44	41.9	898	11	Q8K1S4	Q8K1S4 mus musculu
70	44	41.9	898	11	Q8B721	Q8B721 rattus norv
71	43	41.0	167	10	Q946C3	Q946C3 theobroma c
72	43	41.0	269	2	Q93E62	Q93E62 cowdria rum
73	43	41.0	281	2	Q9S6H1	Q9S6H1 cowdria rum
74	43	41.0	281	2	Q46328	Q46328 cowdria rum
75	43	41.0	281	2	Q9S6H0	Q9S6H0 cowdria rum
76	43	41.0	306	11	Q8CC62	Q8CC62 mus musculu
77	43	41.0	367	2	Q9XD58	Q9XD58 pseudomonas
78	43	41.0	498	17	Q97V40	Q97V40 sulfolobus
79	43	41.0	1127	5	Q8T5I9	Q8T5I9 anopheles g
80	43	41.0	2028	11	Q8C1R4	Q8C1R4 mus musculu
81	43	41.0	5188	16	Q8X4H5	Q8X4H5 escherichia
82	43	41.0	5291	16	Q8X2T1	Q8X2T1 escherichia
83	42.5	40.5	417	10	Q9S1S4	Q9S1S4 arabidopsis
84	42.5	40.5	180	1	Q50522	Q50522 methanobact
85	42	40.0	220	10	Q9LVU1	Q9LVU1 arabidopsis
86	42	40.0	238	10	Q9FPE1	Q9FPE1 arabidopsis
87	42	40.0	244	10	Q8L3T0	Q8L3T0 arabidopsis
88	42	40.0	363	2	Q51860	Q51860 pseudomonas
89	42	40.0	392	16	Q8PK80	Q8PK80 xanthomonas

90 Q8P8p8 xanthomonas
 91 Q13392 homo sapien
 92 Q96C62 homo sapien
 93 Q8WU74 homo sapien
 94 Q9BU23 homo sapien
 95 Q9NKR9 homo sapien
 96 Q9BX84 homo sapien
 97 Q9DJY3 human coxa
 98 Q9H5L6 homo sapien
 99 Q8U120 pyrococcus
 100 Q911J3 pseudomonas

ALIGNMENTS

RESULT 1

ID Q52105 PRELIMINARY; PRT; 286 AA.
 AC Q52105;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE OMP-1D (28kDa outer membrane protein gene 16).
 GN OMP-1D.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OC NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=98084465; PubMed=9423849;
 RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 are encoded by a polymorphic multigene family.";
 RL Infect. Immun. 66:132-139(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=98321180; PubMed=9647746;
 RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
 RA Allemen A.R.;
 RT "Molecular characterization of a 28 kDa surface antigen gene family of
 the tribe Ehrlichiae.";
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Osciola;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; U72291; AAC02938.1;
 DR EMBL; AF479833; AAC26718.1;
 DR EMBL; AF479834; AAC12935.1;
 DR InterPro; IPR002366; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 286 AA; 31509 MW; F145A79270F386BE CRC64;

Query Match 100.0%; Score 105; DB 2; Length 286;
 Best Local Similarity 100.0%; Pred. No. 6.6e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
 |||||
 Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 2

Q8GGU3

ID Q8GGU3 PRELIMINARY; PRT; 287 AA.
 AC Q8GGU3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 28kDa outer membrane protein gene 16.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OC NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=St. Vincent;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479837; AAC12951.1;
 SQ SEQUENCE 287 AA; 31504 MW; A9592249C83695B2 CRC64;

Query Match 100.0%; Score 105; DB 2; Length 287;
 Best Local Similarity 100.0%; Pred. No. 6.6e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
 |||||
 Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 3

ID Q8GGU2 PRELIMINARY; PRT; 291 AA.
 AC Q8GGU2;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 28kDa outer membrane protein gene 16.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OC NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wakulla;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479838; AAC12956.1;
 SQ SEQUENCE 291 AA; 32005 MW; 22D35EDED283195A CRC64;

Query Match 100.0%; Score 105; DB 2; Length 291;
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
 |||||
 Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 4

ID Q8G921 PRELIMINARY; PRT; 291 AA.
 AC Q8G921;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE 28kDa outer membrane protein gene 16.
 OS Ehrlichia chaffeensis.

```
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=West Paces, and Heartland;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479833; AAO12941.1; -.
DR EMBL; AF479836; AAO12946.1; -.
SQ SEQUENCE 291 AA; 31931 MW; 22D35A2202831369 CRC64;

Query Match 100.0%; Score 105; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 6.7e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 5
Q8G8P3
ID Q8G8P3 PRELIMINARY; PRT; 291 AA.
AC Q8G8P3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479839; AAO12962.1; -.
DR EMBL; AF479840; AAO12968.1; -.
SQ SEQUENCE 291 AA; 32130 MW; 8220C436A1FC08E2 CRC64;

Query Match 95.2%; Score 100; DB 2; Length 291;
Best Local Similarity 94.7%; Pred. No. 4.4e-08;
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 6
O52107
ID O52107 PRELIMINARY; PRT; 280 AA.
AC O52107;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Major outer membrane protein OMP-1F (28kDa outer membrane protein gene
18).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
Membrane Protein Multigene Family in Ehrlichia canis and E.
chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burrig M.J.,
Allen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02940.1; -.
DR EMBL; AF479833; AAO12931.1; -.
DR EMBL; AF479834; AAO12937.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30731 MW; CCAA6C34E2AF393E CRC64;

Query Match 70.5%; Score 74; DB 2; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00071;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 7
Q93DD2
ID Q93DD2 PRELIMINARY; PRT; 281 AA.
AC Q93DD2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V6;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393392; AAL12922.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
```



```
DE Outer membrane protein p28 (Fragment).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St.Vincent;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL; AF077735; AAC31548.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 66.7%; Score 70; DB 2; Length 246;
Best Local Similarity 66.7%; Pred. No. 0.0027;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWRDRCVI 18
   ||| ||||:|||| |
Db 29 NTTAGVFLKQDWDGSAI 46

RESULT 11
Q93DD4 ID Q93DD4 PRELIMINARY; PRT; 275 AA.
AC Q93DD4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V2;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393389; AAL12919.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 29974 MW; 2ECCF2F988B2E9D9 CRC64;

Query Match 66.7%; Score 70; DB 2; Length 275;
Best Local Similarity 66.7%; Pred. No. 0.0031;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWRDRCVI 18
   ||| ||||:|||| |
Db 59 NTTAGVFLKQDWDGSAI 76

RESULT 12
Q93DD1 ID Q93DD1 PRELIMINARY; PRT; 276 AA.
AC Q93DD1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393393; AAL12923.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30028 MW; 2D7143AFCBFF2EBE CRC64;

Query Match 66.7%; Score 70; DB 2; Length 276;
Best Local Similarity 66.7%; Pred. No. 0.0031;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWRDRCVI 18
   ||| ||||:|||| |
Db 59 NTTAGVFLKQDWDGSAI 76

RESULT 13
O85817 ID O85817 PRELIMINARY; PRT; 276 AA.
AC O85817;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28 kDa outer membrane protein).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V9;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Chattanooga, West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077734; AAC31547.1; -
DR EMBL; AF393395; AAL12925.1; -
DR EMBL; AY117397; AAM77032.1; -
DR EMBL; AF479835; AAO12943.1; -
DR EMBL; AF479836; AAO12948.1; -
DR EMBL; AF479837; AAO12953.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 66.7%; Score 70; DB 2; Length 276;
```

Best Local Similarity 66.7%; Pred. No. 0.0031;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWRVCVI 18
DB 59 NTTAGVFLKQDWDGSAI 76

RESULT 14

Q8GGUO PRELIMINARY; PRT; 276 AA.
AC Q8GGUO;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE 28kDa outer membrane protein gene 19.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479838; AAC12958.1; -;
SQ SEQUENCE 276 AA; 30045 MW; 8E3C9719D3C67A64 CRC64;

Query Match 66.7%; Score 70; DB 2; Length 276;
Best Local Similarity 66.7%; Pred. No. 0.0031;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWRVCVI 18
DB 59 NTTAGVFLKQDWDGSAI 76

RESULT 15

Q93DD3 PRELIMINARY; PRT; 280 AA.
AC Q93DD3;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V5;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393391; AAL12921.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30372 MW; C7B8C8710BC167E9 CRC64;

Query Match 65.7%; Score 69; DB 2; Length 280;
Best Local Similarity 63.2%; Pred. No. 0.0046;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWRVCVI 19
DB 59 STTAGVFLKQDWDGSAIS 77

RESULT 16

Q9ZGM9 PRELIMINARY; PRT; 280 AA.
AC Q9ZGM9;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077733; AAC31546.1; -;
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Query Match 65.7%; Score 69; DB 2; Length 280;
Best Local Similarity 63.2%; Pred. No. 0.0046;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWRVCVI 19
DB 59 STTAGVFLKQDWDGSAIS 77

RESULT 17

O85816 PRELIMINARY; PRT; 280 AA.
AC O85816;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28kDa outer membrane protein gene
DE 19).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V8, and V4;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

```

[4]
RN  SEQUENCE FROM N.A.
RP  STRAIN=Liberty, and Jax;
RX  PubMed=12496165;
RA  Cheng C., Paddock C.D., Ganta R.R.;
RT  "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT  by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
RT  and Other Regions of the Genome.";
RL  Infect. Immun. 71:187-195(2003).
DR  EMBL: AF077732; AAC31545.1; -
DR  EMBL: AF393394; AAL12924.1; -
DR  EMBL: AF393390; AAL12920.1; -
DR  EMBL: AF479839; AAO12964.1; -
DR  EMBL: AF479840; AAO12970.1; -
DR  InterPro: IPR002566; Surface_Ag_msp4.
DR  Pfam: PF01617; Surface_Ag_2; 1.
SQ  SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match      65.7%; Score 69; DB 2; Length 280;
Best Local Similarity 63.2%; Pred. No. 0.0046;
Matches 12; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY  1 NTTGVGFIQDWDRCVIS 19
    :||| |||||:|||| ||
DB  59 STTAGVGLKQDWDGSAIS 77

RESULT 18
Q9ADV3  PRELIMINARY; PRT; 280 AA.
AC  Q9ADV3;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE  Major outer membrane protein P30-2.
GN  P30-2.
OS  Ehrlichia canis.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Anaplasmataceae; Ehrlichia.
OX  NCBI_TaxID=944;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Oklahoma;
RX  MEDLINE=98371112; PubMed=9705412;
RA  Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT  "Cloning and characterization of multigenes encoding the
RT  immunodominant 30-kilodalton major outer membrane proteins of
RT  Ehrlichia canis and application of the recombinant protein for
RT  serodiagnosis.";
RL  J. Clin. Microbiol. 36:2671-2680(1998).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Oklahoma;
RX  MEDLINE=21153566; PubMed=11254561;
RA  Ohashi N., Rikihisa Y., Unver A.;
RT  "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT  Membrane Protein Multigene Family in Ehrlichia canis and E.
RT  chaffeensis.";
RL  Infect. Immun. 69:2083-2091(2001).
DR  EMBL: AF078553; AAK28699.1; -
DR  InterPro: IPR002566; Surface_Ag_msp4.
DR  Pfam: PF01617; Surface_Ag_2; 1.
SQ  SEQUENCE 280 AA; 30803 MW; 27238BE1C7E68A91 CRC64;

Query Match      59.0%; Score 62; DB 2; Length 280;
Best Local Similarity 57.9%; Pred. No. 0.063;
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY  1 NTTGVGFIQDWDRCVIS 19
    :||| |||||:|||| ||
DB  59 NSTAGVGLKHDWNGTIS 77

RESULT 19
Q9F472  PRELIMINARY; PRT; 280 AA.
ID  Q9F472;
AC  Q9F472;
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  P28-7.
GN  P28-7.
OS  Ehrlichia canis.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Anaplasmataceae; Ehrlichia.
OX  NCBI_TaxID=944;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Jake;
RX  MEDLINE=99242757; PubMed=10225842;
RA  McBride J.W., Yu, Xj, Walker D.H.;
RT  "Molecular cloning of the gene for a conserved major immunoreactive
RT  28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT  antigen.";
RL  Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Jake;
RX  MEDLINE=20432107; PubMed=10974556;
RA  McBride J.W., Yu X.J., Walker D.H.;
RT  "A conserved, transcriptionally active p28 multigene locus of
RT  Ehrlichia canis.";
RL  Gene 254:245-252(2000).
DR  EMBL: AF082744; AAG14361.1; -
DR  InterPro: IPR002566; Surface_Ag_msp4.
DR  Pfam: PF01617; Surface_Ag_2; 1.
SQ  SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match      59.0%; Score 62; DB 2; Length 280;
Best Local Similarity 57.9%; Pred. No. 0.063;
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY  1 NTTGVGFIQDWDRCVIS 19
    :||| |||||:|||| ||
DB  59 NSTAGVGLKHDWNGTIS 77

RESULT 20
Q9F472  PRELIMINARY; PRT; 278 AA.
ID  Q9F472;
AC  Q9F472;
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE  P28-7.
GN  P28-7.
OS  Ehrlichia canis.
OC  Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC  Anaplasmataceae; Ehrlichia.
OX  NCBI_TaxID=944;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Jake;
RX  MEDLINE=99242757; PubMed=10225842;
RA  McBride J.W., Yu, Xj, Walker D.H.;
RT  "Molecular cloning of the gene for a conserved major immunoreactive
RT  28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT  antigen.";
RL  Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Jake;
RX  MEDLINE=20432107; PubMed=10974556;
RA  McBride J.W., Yu X.J., Walker D.H.;
RT  "A conserved, transcriptionally active p28 multigene locus of
RT  Ehrlichia canis.";

```

```

RL Gene 254:245-252(2000).
DR EMBL: AF082744; AAC64550.2; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30485 MW; 2411CAAB4C56CA74 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVGFGIEQDWD 14
Db :|||||:|
61 STVGVEGLKHDWD 73

RESULT 21
Q9R8A8
ID Q9R8A8 PRELIMINARY; PRT; 278 AA.
AC Q9R8A8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082746; AAC64552.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVGFGIEQDWD 14
Db :|||||:|
61 STVGVEGLKHDWD 73

RESULT 22
Q9R8A7
ID Q9R8A7 PRELIMINARY; PRT; 278 AA.
AC Q9R8A7;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Emon;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082747; AAC64553.1; -.

```

```

DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVGFGIEQDWD 14
Db :|||||:|
61 STVGVEGLKHDWD 73

RESULT 23
Q9R3J3
ID Q9R3J3 PRELIMINARY; PRT; 278 AA.
AC Q9R3J3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Florida;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082750; AAC64556.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVGFGIEQDWD 14
Db :|||||:|
61 STVGVEGLKHDWD 73

RESULT 24
Q9R8A6
ID Q9R8A6 PRELIMINARY; PRT; 278 AA.
AC Q9R8A6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DJ;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082748; AAC64554.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.

```

DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTVGVFGIEQDWD 14
:|||||::|||
Db 61 STVGVEGLKHDWD 73

RESULT 25

Q9R8A9
ID Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Louisiana;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, XJ, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082745; AAC64551.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTVGVFGIEQDWD 14
:|||||::|||
Db 61 STVGVEGLKHDWD 73

RESULT 26

Q9R8A5
ID Q9R8A5 PRELIMINARY; PRT; 278 AA.
AC Q9R8A5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fuzzy;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, XJ, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082749; AAC64555.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.

FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.28;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTVGVFGIEQDWD 14
:|||||::|||
Db 61 STVGVEGLKHDWD 73

RESULT 27

Q9ZGJ1
ID Q9ZGJ1 PRELIMINARY; PRT; 307 AA.
AC Q9ZGJ1;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Major outer membrane protein P30-1.
GN P30-1.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT Chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL: AF078553; AAC68666.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 307 AA; 34108 MW; 9DA9FD63EBF8EC97 CRC64;

Query Match 55.2%; Score 58; DB 2; Length 307;
Best Local Similarity 69.2%; Pred. No. 0.31;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTVGVFGIEQDWD 14
:|||||::|||
Db 90 STVGVEGLKHDWD 102

RESULT 28

Q9F475
ID Q9F475 PRELIMINARY; PRT; 276 AA.
AC Q9F475;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE P28-3 (Major outer membrane protein P30-4).
GN P28-3 OR P30-4.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Oklahoma;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-Oklahoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT Chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF082744; AAG14358.1; -
DR EMBL; AF324792; AAK31313.1; -
DR EMBL; AF078553; AAK28697.1; -
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30659 MW; CE51AB37D17AF3A4 CRC64;

Query Match 51.4%; Score 54; DB 2; Length 276;
Best Local Similarity 44.4%; Pred. No. 1.2;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDMDRCVI 18
Db 60 NTTGIFGLKESWTGGII 77

RESULT 29
Q8G948 PRELIMINARY; PRT; 276 AA.
AC Q8G948
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 28kDa outer membrane protein gene 15.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-West Paces, Heartland, St. Vincent, and Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479835; AA012940.1; -
DR EMBL; AF479836; AA012941.1; -
DR EMBL; AF479837; AA012952.1; -
SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

DR EMBL; AF479837; AA012950.1; -
DR EMBL; AF479838; AA012955.1; -
SQ SEQUENCE 276 AA; 30316 MW; 0D6F5353F9C0F17C CRC64;

Query Match 48.6%; Score 51; DB 2; Length 276;
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDMD 14
Db 60 NATVALYGLKQDN 73

RESULT 30
Q8GGU1 PRELIMINARY; PRT; 277 AA.
ID Q8GGU1
AC Q8GGU1
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479835; AA012942.1; -
DR EMBL; AF479836; AA012947.1; -
DR EMBL; AF479837; AA012952.1; -
SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

Query Match 48.6%; Score 51; DB 2; Length 277;
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDMD 14
Db 60 NATVALYGLKQDN 73

RESULT 31
Q8GW7 PRELIMINARY; PRT; 277 AA.
ID Q8GW7
AC Q8GW7
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479835; AA012942.1; -
DR EMBL; AF479836; AA012947.1; -
DR EMBL; AF479837; AA012952.1; -
SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

```
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NTTVGVFGIEQDWD 14
Db 60 NATVALYGLKODWN 73

RESULT 32
O52106 PRELIMINARY; PRT; 278 AA.
AC O52106;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE OMP-1E (28kDa outer membrane protein gene 17).
GN OMP-1E
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikhiisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02939.1; -.
DR EMBL; AF479833; AAC02936.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30543 MW; E321E3CA259B87FD CRC64;

Query Match 48.6%; Score 51; DB 2; Length 278;
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NTTVGVFGIEQDWD 14
Db 60 NPTVALYGLKODWE 73

RESULT 33
O8G8Q5 PRELIMINARY; PRT; 278 AA.
AC O8G8Q5;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
```

```
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479833; AAC02963.1; -.
DR EMBL; AF479840; AAC02969.1; -.
SQ SEQUENCE 278 AA; 30381 MW; 7626778E5798D6B7 CRC64;

Query Match 48.6%; Score 51; DB 2; Length 278;
Best Local Similarity 50.0%; Pred. No. 3.8;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NTTVGVFGIEQDWD 14
Db 60 NPTVALYGLKODWE 73

RESULT 34
O52104 PRELIMINARY; PRT; 280 AA.
AC O52104; O85357;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE OMP-1C (28kDa outer membrane protein 15).
GN OMP-1C.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikhiisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,
RA Alleman A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=22384137; PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02937.1; -.
DR EMBL; AF479833; AAC02930.1; -.
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30322 MW; BD835E491086DF01 CRC64;

Query Match 47.6%; Score 50; DB 2; Length 280;
Best Local Similarity 50.0%; Pred. No. 5.6;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```

```

QY      1 NNTVGVFQIEQDWD 14
DB      60 NPTVALYGLKQDWN 73

RESULT 35
ID      O8GGU5          PRELIMINARY;          PRT;      280 AA.
AC      O1GGU5;
DT      01-03-2003 (TEMBLrel.. 23, Created)
DT      01-WAR-2003 (TEMBLrel.. 23, Last sequence update)
DT      01-WAR-2003 (TEMBLrel.. 23, Last annotation update)
DE      28DA outer membrane protein gene 15.
OS      Ehrlichia chaffeensis.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC      Anaplasmataceae; Ehrlichia.
OX      NCBI_TaxId=945;
      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Oscila.
FX      PubMed=12496165;
RA      Cheng C., Paddock C.D., Ganta R.R.;
RT      "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT      by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT      and Other Regions of the Genome. ";
RL      Infect. Immun. 71:167-195 (2003).
SQ      EMBL: AF479834; AAC12934.1.
SQ      SEQUENCE      280 AA; 30332 MW;  BD835D792386DF01  CRC64;

Query Match      47.68;      Score 50; DB 2; Length 280;
Best Local Similarity 50.08; Pred: No. 5.6;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 NNTVGVFQIEQDWD 14
DB      60 NPTVALYGLKQDWN 73

Search completed: October 6, 2003, 07:52:39
Job time : 28.806 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 11.0597 Seconds
(without alignments)
72.688 Million cell updates/sec

Title: US-09-765-739A-5

Perfect score: 105

Sequence: 1 NNTVGVEIQDWDRCVIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Issued Patents_AA: *
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/iaa/PTDUS_COMB.pep: *
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	286	3	US-08-953-326-15
2	105	100.0	286	4	US-09-660-587-12
3	105	100.0	286	4	US-09-261-358A-12
4	105	100.0	286	4	US-09-201-458-8
5	105	100.0	286	4	US-09-314-701-8
6	77	73.3	280	3	US-08-953-326-17
7	74	70.5	280	4	US-09-660-587-14
8	74	70.5	280	4	US-09-261-358A-14
9	74	70.5	280	4	US-09-201-458-10
10	74	70.5	280	4	US-09-314-701-12
11	73	69.5	276	3	US-08-953-326-18
12	73	69.5	280	3	US-08-733-230-4
13	73	69.5	280	3	US-08-953-326-4
14	73	69.5	281	4	US-09-660-587-9
15	73	69.5	281	4	US-09-261-358A-9
16	73	69.5	281	4	US-09-201-458-5
17	73	69.5	281	4	US-09-314-701-2
18	71	67.6	288	4	US-09-314-701-32
19	62	59.0	280	4	US-09-660-587-6
20	62	59.0	280	4	US-09-261-358A-6
21	62	59.0	280	4	US-09-314-701-38
22	58	55.2	278	4	US-09-660-587-2
23	58	55.2	278	4	US-09-261-358A-2
24	58	55.2	278	4	US-09-201-458-2
25	58	55.2	307	4	US-09-314-701-36
26	54	51.4	276	4	US-09-660-587-44
27	54	51.4	276	4	US-09-314-701-42
28	48.6				Sequence 15, Appl
29	48.6				Sequence 12, Appl
30	48.6				Sequence 12, Appl
31	48.6				Sequence 8, Appl
32	48.6				Sequence 8, Appl
33	48.6				Sequence 17, Appl
34	48.6				Sequence 14, Appl
35	48.6				Sequence 14, Appl
36	48.6				Sequence 10, Appl
37	48.6				Sequence 18, Appl
38	48.6				Sequence 4, Appl
39	48.6				Sequence 4, Appl
40	48.6				Sequence 4, Appl
41	48.6				Sequence 4, Appl
42	48.6				Sequence 4, Appl
43	48.6				Sequence 4, Appl
44	48.6				Sequence 4, Appl
45	48.6				Sequence 4, Appl
46	48.6				Sequence 4, Appl
47	48.6				Sequence 4, Appl
48	48.6				Sequence 4, Appl
49	48.6				Sequence 4, Appl
50	48.6				Sequence 4, Appl
51	48.6				Sequence 4, Appl
52	48.6				Sequence 4, Appl
53	48.6				Sequence 4, Appl
54	48.6				Sequence 4, Appl
55	48.6				Sequence 4, Appl
56	48.6				Sequence 4, Appl
57	48.6				Sequence 4, Appl
58	48.6				Sequence 4, Appl
59	48.6				Sequence 4, Appl
60	48.6				Sequence 4, Appl
61	48.6				Sequence 4, Appl
62	48.6				Sequence 4, Appl
63	48.6				Sequence 4, Appl
64	48.6				Sequence 4, Appl
65	48.6				Sequence 4, Appl
66	48.6				Sequence 4, Appl
67	48.6				Sequence 4, Appl
68	48.6				Sequence 4, Appl
69	48.6				Sequence 4, Appl
70	48.6				Sequence 4, Appl
71	48.6				Sequence 4, Appl
72	48.6				Sequence 4, Appl
73	48.6				Sequence 4, Appl
74	48.6				Sequence 4, Appl
75	48.6				Sequence 4, Appl
76	48.6				Sequence 4, Appl
77	48.6				Sequence 4, Appl
78	48.6				Sequence 4, Appl
79	48.6				Sequence 4, Appl
80	48.6				Sequence 4, Appl
81	48.6				Sequence 4, Appl
82	48.6				Sequence 4, Appl
83	48.6				Sequence 4, Appl
84	48.6				Sequence 4, Appl
85	48.6				Sequence 4, Appl
86	48.6				Sequence 4, Appl
87	48.6				Sequence 4, Appl
88	48.6				Sequence 4, Appl
89	48.6				Sequence 4, Appl
90	48.6				Sequence 4, Appl
91	48.6				Sequence 4, Appl
92	48.6				Sequence 4, Appl
93	48.6				Sequence 4, Appl
94	48.6				Sequence 4, Appl
95	48.6				Sequence 4, Appl
96	48.6				Sequence 4, Appl
97	48.6				Sequence 4, Appl
98	48.6				Sequence 4, Appl
99	48.6				Sequence 4, Appl
100	48.6				Sequence 4, Appl

Sequence 16, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 9, Appl
Sequence 10, Appl
Sequence 11, Appl
Sequence 7, Appl
Sequence 6, Appl
Sequence 15, Appl
Sequence 15, Appl
Sequence 11, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 25374, A
Sequence 25374, A
Sequence 21412, A
Sequence 19757, A
Sequence 6800, Ap
Sequence 9, Appl
Sequence 9, Appl
Sequence 16, Appl
Sequence 17, Appl
Sequence 24741, A
Sequence 7, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 21630, A
Sequence 20, Appl
Sequence 6900, Ap
Sequence 17, Appl
Sequence 20, Appl
Sequence 26, Appl
Sequence 29334, A
Sequence 28754, A
Sequence 3, Appl
Sequence 225, App
Sequence 8, Appl
Sequence 7, Appl
Sequence 4333, Ap
Sequence 48, Appl
Sequence 49, Appl
Sequence 16, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 1, Appl
Sequence 22, Appl
Sequence 367, App
Sequence 25318, A
Sequence 25490, A
Sequence 5, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 732, App
Sequence 469, App
Sequence 6, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 6212, Ap
Sequence 2, Appl
Sequence 2, Appl
Sequence 18, Appl
Sequence 28143, A
Sequence 4, Appl
Sequence 14, Appl
Sequence 20332, A
Sequence 31546, A
Sequence 414, App
Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Gaeta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 100.0%; Score 105; DB 3; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||

RESULT 2
US-09-660-587-12
; Sequence 12, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-660-587-12

Query Match 100.0%; Score 105; DB 4; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||

Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 3
US-09-261-358A-12
; Sequence 12, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-261-358A-12

Query Match 100.0%; Score 105; DB 4; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||

RESULT 4
US-09-201-458-8
; Sequence 8, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-201-458-8

Query Match 100.0%; Score 105; DB 4; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||

RESULT 5
US-09-314-701-8
; Sequence 8, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko

```

; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-660-587-14

Query Match 70.5%; Score 74; DB 4; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00031;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVGFIEQDWDRCVIS 19
    ||| ||||:|||| ||
Db 60 NTTGVFGLKQDWDGSTIS 78

RESULT 8
US-09-261-358A-14
; Sequence 14, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-261-358A-14

Query Match 70.5%; Score 74; DB 4; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00031;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVGFIEQDWDRCVIS 19
    ||| ||||:|||| ||
Db 60 NTTGVFGLKQDWDGSTIS 78

RESULT 9
US-09-201-458-10
; Sequence 10, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 10
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:

```

; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-201-458-10

Query Match 70.5%; Score 74; DB 4; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00031;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVGIEQDWDRCVIS 19
Db 60 NTTGVGGLKQNDGSGTIS 78

RESULT 10
US-09-314-701-12
; Sequence 12, Application US/09314701
; Patent No. 6544317
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. 654451710
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-12

Query Match 70.5%; Score 74; DB 4; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00031;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVGIEQDWDRCVIS 19
Db 60 NTTGVGGLKQNDGSGTIS 78

RESULT 11
US-08-953-326-18
; Sequence 19, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match 69.5%; Score 73; DB 3; Length 276;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGIEQDWDRCVIS 19
Db 59 NTTGVGGLKQNDGSGAIS 77

RESULT 12
US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/08/733,230
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; SEQ ID NO 4
; LENGTH: 280
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-4

Query Match 69.5%; Score 73; DB 3; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGIEQDWDRCVIS 19
Db 60 NTTGVGGLKQNDGSGAIS 78

RESULT 13
US-08-953-326-4
; Sequence 4, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
US-08-953-326-4

```
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match          69.5%; Score 73; DB 3; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
   |||||:::|:| ||
Db 60 NTTGVFGLKQNWGSAIS 78

RESULT 14
US-09-660-587-9
; Sequence 9, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-660-587-9

Query Match          69.5%; Score 73; DB 4; Length 281;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
   |||||:::|:| ||
Db 59 NTTGVFGLKQNWGSAIS 77

RESULT 15
US-09-261-358A-9
; Sequence 9, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
```

```
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-261-358A-9

Query Match          69.5%; Score 73; DB 4; Length 281;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
   |||||:::|:| ||
Db 59 NTTGVFGLKQNWGSAIS 77

RESULT 16
US-09-201-458-5
; Sequence 5, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immureactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-201-458-5

Query Match          69.5%; Score 73; DB 4; Length 281;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
   |||||:::|:| ||
Db 59 NTTGVFGLKQNWGSAIS 77

RESULT 17
US-09-314-701-2
; Sequence 2, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-2

Query Match          69.5%; Score 73; DB 4; Length 281;
Best Local Similarity 68.4%; Pred. No. 0.00044;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
; QY 1 NTTGVFGIEQDWDRCVTS 19
;      |||||:::|:| |
; Db 59 NTTGVFGLKQWDGSAIS 77
;
; RESULT 18
; US-09-314-701-32
; Sequence 32, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314.701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; US-09-314-701-32
;
; Query Match 67.6%; Score 71; DB 4; Length 288;
; Best Local Similarity 66.7%; Pred. No. 0.0095;
; Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 NTTGVFGIEQDWDRCVI 18
;      |||||:::|:| |
; Db 60 NTTGVFGLKQWDGATI 77
;
; RESULT 19
; US-09-660-587-6
; Sequence 6, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660.587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261.358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-6 protein
; US-09-660-587-6
;
; Query Match 59.0%; Score 62; DB 4; Length 280;
; Best Local Similarity 57.9%; Pred. No. 0.026;
; Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 NTTGVFGIEQDWDRCVTS 19
;      |||||:::|:| |
; Db 59 NSTGVFGLKHDWNGGTIS 77
;
; RESULT 20
; US-09-261-358A-6
; Sequence 6, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
```

```
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/261.358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201.458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28SA3 protein
; US-09-261-358A-6
;
; Query Match 59.0%; Score 62; DB 4; Length 280;
; Best Local Similarity 57.9%; Pred. No. 0.026;
; Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 NTTGVFGIEQDWDRCVIS 19
;      |||||:::|:| |
; Db 59 NSTGVFGLKHDWNGGTIS 77
;
; RESULT 21
; US-09-314-701-38
; Sequence 38, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314.701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; US-09-314-701-38
;
; Query Match 59.0%; Score 62; DB 4; Length 280;
; Best Local Similarity 57.9%; Pred. No. 0.026;
; Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 NTTGVFGIEQDWDRCVIS 19
;      |||||:::|:| |
; Db 59 NSTGVFGLKHDWNGGTIS 77
;
; RESULT 22
; US-09-660-587-2
; Sequence 2, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660.587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261.358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
```

```
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein
US-09-660-587-2

Query Match          55.2%; Score 58; DB 4; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.11;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWD 14
   :|||||::|||
Db 61 STVGFGKLKHDWD 73

RESULT 23
US-09-261-358A-2
; Sequence 2, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28-1 protein
US-09-261-358A-2

Query Match          55.2%; Score 58; DB 4; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.11;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWD 14
   :|||||::|||
Db 61 STVGFGKLKHDWD 73

RESULT 24
US-09-201-458-2
; Sequence 2, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a 30 kDa immunoreactive
US-09-201-458-2
```

```
Query Match          55.2%; Score 58; DB 4; Length 278;
Best Local Similarity 69.2%; Pred. No. 0.11;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWD 14
   :|||||::|||
Db 61 STVGFGKLKHDWD 73

RESULT 25
US-09-314-701-36
; Sequence 36, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-36

Query Match          55.2%; Score 58; DB 4; Length 307;
Best Local Similarity 69.2%; Pred. No. 0.12;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTGVFGIEQDWD 14
   :|||||::|||
Db 90 STVGFGKLKHDWD 102

RESULT 26
US-09-660-587-44
; Sequence 44, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 44
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-3 protein
US-09-660-587-44

Query Match          51.4%; Score 54; DB 4; Length 276;
Best Local Similarity 44.4%; Pred. No. 0.48;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTVGFIEQDWDRCVI 18
   ||| |::|::|
Db 60 NTTTGIFGLKESWTGGII 77

RESULT 27
US-09-314-701-42
```

; Sequence 42, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314.701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-42

Query Match 51.48; Score 54; DB 4; Length 276;
Best Local Similarity 44.48; Pred. No. 0.48;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
QY 1 NTTGVGFGIEQDWDRCVI 18
Db 60 NTTGIFGLKESWTGGII 77

RESULT 28
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953.326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 48.68; Score 51; DB 3; Length 278;
Best Local Similarity 50.08; Pred. No. 1.5;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTGVGFGIEQDWD 14
Db 60 NPTVALYGLKQDWE 73

RESULT 29
US-09-660-587-13
; Sequence 13, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie

; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660.587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261.358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-660-587-13

Query Match 48.68; Score 51; DB 4; Length 278;
Best Local Similarity 50.08; Pred. No. 1.5;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTGVGFGIEQDWD 14
Db 60 NPTVALYGLKQDWE 73

RESULT 30
US-09-261-358A-13
; Sequence 13, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261.358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201.458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-261-358A-13

Query Match 48.68; Score 51; DB 4; Length 278;
Best Local Similarity 50.08; Pred. No. 1.5;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 NTTGVGFGIEQDWD 14
Db 60 NPTVALYGLKQDWE 73

RESULT 31
US-09-201-458-9
; Sequence 9, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201.458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9

```
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-201-458-9

Query Match      48.6%; Score 51; DB 4; Length 278;
Best Local Similarity 50.0%; Pred. No. 1.5;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWD 14
Db 60 NPTVALYGLKQDWN 73

RESULT 32
US-09-314-701-10
; Sequence 10, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-10

Query Match      48.6%; Score 51; DB 4; Length 278;
Best Local Similarity 50.0%; Pred. No. 1.5;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWD 14
Db 60 NPTVALYGLKQDWE 73

RESULT 33
US-09-660-587-11
; Sequence 11, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-660-587-11

Query Match      47.6%; Score 50; DB 4; Length 280;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 NTTGVFGIEQDWD 14
Db 60 NPTVALYGLKQDWN 73

RESULT 34
US-09-261-358A-11
; Sequence 11, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-261-358A-11

Query Match      47.6%; Score 50; DB 4; Length 280;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWD 14
Db 60 NPTVALYGLKQDWN 73

RESULT 35
US-09-201-458-7
; Sequence 7, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 7
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-201-458-7

Query Match      47.6%; Score 50; DB 4; Length 280;
Best Local Similarity 50.0%; Pred. No. 2.1;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWD 14
Db 60 NPTVALYGLKQDWN 73

Search completed: October 6, 2003, 07:53:54
Job time : 13.0597 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:03 ; Search time 33.6045 Seconds
(without alignments)
89.744 Million cell updates/sec

Title: US-09-765-739A-5
Perfect score: 105
Sequence: 1 NTTCVGFGEQDWDRCVIS 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	105	100.0	19	ABG30747
2	105	100.0	286	19 AAW51092
3	105	100.0	286	20 AAY06946
4	105	100.0	286	21 AAB36186
5	105	100.0	286	22 AAU04196
6	105	100.0	286	23 ABG77938
7	105	100.0	286	23 AAU96108
8	105	100.0	286	23 AAU73415
9	77	73.3	280	19 AAW51094

10	77	73.3	280	21	AAAB36188	Ehrlichia chaffeen
11	77	73.3	280	22	AAU04198	Variable surface a
12	77	73.3	280	23	AAU73417	Ehrlichia chaffeen
13	77	73.3	280	23	ABG30749	Ehrlichia chaffeen
14	74	70.5	280	19	AAAY06948	E. chafeensis OMP-
15	74	70.5	280	23	ABG77940	Ehrlichia chaffeen
16	74	70.5	280	23	AAU96110	Ehrlichia chafeens
17	74	70.5	280	23	ABG30745	Ehrlichia chaffeen
18	73	69.5	256	20	AAAY05942	E. chafeensis p28
19	73	69.5	256	23	ABG77966	Protein encoded by
20	73	69.5	276	19	AAW51095	Ehrlichia chaffeen
21	73	69.5	276	21	AAAB36189	Ehrlichia chaffeen
22	73	69.5	276	22	AAU04199	Variable surface a
23	73	69.5	280	19	AAW51089	Ehrlichia chaffeen
24	73	69.5	280	21	AAAB36183	Ehrlichia chaffeen
25	73	69.5	280	22	AAU04193	Major antigenic pr
26	73	69.5	281	20	AAAY06943	E. chafeensis OMP-
27	73	69.5	281	23	ABG77935	Ehrlichia chaffeen
28	73	69.5	281	23	AAU96105	Ehrlichia chafeens
29	73	69.5	281	23	AAU73418	Ehrlichia chaffeen
30	71	67.6	20	23	ABG30744	Ehrlichia canis pe
31	71	67.6	288	20	AAAY05959	E. canis p30 prote
32	71	67.6	288	23	ABG77950	Ehrlichia canis ou
33	62	59.0	280	20	AAAY06962	E. canis p30-2 pro
34	62	59.0	280	21	AAAY71479	Ehrlichia canis im
35	62	59.0	280	23	ABG77953	Ehrlichia canis ou
36	62	59.0	280	23	AAU96102	Ehrlichia canis p2
37	58	55.2	20	23	ABG30743	Ehrlichia canis pe
38	58	55.2	278	21	AAAY71477	Ehrlichia canis p2
39	58	55.2	278	23	AAU96100	Ehrlichia canis im
40	58	55.2	307	20	AAAY05961	E. canis p30-1 p30
41	58	55.2	307	23	ABG77952	Ehrlichia canis ou
42	54	51.4	276	20	AAAY05964	Ehrlichia canis ou
43	54	51.4	276	23	AAU96117	Ehrlichia canis p2
44	54	51.4	276	23	AAU96117	Ehrlichia chaffeen
45	51	48.6	18	23	ABG30748	Ehrlichia chaffeen
46	51	48.6	278	19	AAW51093	E. chafeensis OMP-
47	51	48.6	278	20	AAAY05947	Ehrlichia chaffeen
48	51	48.6	278	21	AAAB36187	Ehrlichia chaffeen
49	51	48.6	278	22	AAU04197	Variable surface a
50	51	48.6	278	23	AAE31090	Ehrlichia ruminant
51	51	48.6	278	23	ABG77939	Ehrlichia chaffeen
52	51	48.6	278	23	AAU96109	Ehrlichia chafeens
53	51	48.6	278	23	AAU73416	Ehrlichia chaffeen
54	50	47.6	18	23	ABG30746	Ehrlichia chaffeen
55	50	47.6	280	20	AAAY05945	E. chafeensis OMP-
56	50	47.6	280	23	ABG77937	Ehrlichia chaffeen
57	50	47.6	280	23	AAU96107	Ehrlichia chafeens
58	50	47.6	280	23	AAU73414	Ehrlichia chaffeen
59	46	43.8	65	22	AAU63107	Propionibacterium
60	46	43.8	284	23	AAU96111	Cowdria ruminantiu
61	44	41.9	204	22	ABH11429	Human transmembran
62	44	41.9	204	22	AAW80112	Human protein SEQ
63	44	41.9	287	19	AAW51088	Cowdria ruminantiu
64	44	41.9	287	21	AAAB36182	Cowdria ruminantiu
65	44	41.9	287	22	AAU04192	Major antigenic pr
66	44	41.9	842	23	AAU74818	Human REPR 1 prot
67	44	41.9	898	19	AAW78898	Rat UNC-5 homologu
68	44	41.9	898	23	AAU97899	Human netrin bindi
69	44	41.9	898	23	AAU97900	Rat netrin binding
70	44	41.9	898	23	AAU85403	Human protein NOV1
71	44	41.9	898	23	AAU10543	Rat netrin recepto
72	44	41.9	899	23	AAU79939	Human UNC5-like pr
73	44	41.9	943	22	AAW79128	Human protein SEQ
74	43	41.0	988	22	AAU31162	Human uromodulin S
75	43	41.0	1246	24	ABR98302	Arabidopsis thalia
76	42	40.0	198	21	AAAG07319	Arabidopsis thalia
77	42	40.0	198	21	AAAG38298	Arabidopsis thalia
78	42	40.0	220	21	AAAG08266	Arabidopsis thalia
79	42	40.0	220	21	AAAG14886	Arabidopsis thalia
80	42	40.0	220	21	AAAG48818	Arabidopsis thalia
81	42	40.0	220	21	AAAG48858	Arabidopsis thalia
82	42	40.0	244	21	AAAG07318	Arabidopsis thalia

83 Arabidopsis thaliana
 84 Human polypeptide
 85 Human histidyl-trn
 86 Human TRICH SEQ ID
 87 Novel human protein
 88 Transient receptor
 89 Transient receptor
 90 Human TRICH-6 prot
 91 Vp1/2A protein fra
 92 Human polypeptide
 93 Human nucleic acid
 94 Listeria monocytog
 95 Sweet potato feath
 96 Human multiple reg
 97 Drosophila melanog
 98 Human 5' EST seque
 99 S. pneumoniae deri
 100 Human polypeptide

ALIGNMENTS

RESULT 1
 ABG30747
 ID ABG30747 standard; Peptide; 19 AA.

XX AC ABG30747;
 XX DT 21-OCT-2002 (first entry)
 XX DE Ehrlichia chaffeensis peptide fragment #3.
 XX KW Antibody detection; monoclonal antibody; polyclonal antibody.
 XX OS Ehrlichia chaffeensis.
 XX PN WO200257794-A2.
 XX PD 25-JUL-2002.

XX PF 16-JAN-2002; 2002WO-US01395.

XX PR 18-JAN-2001; 2001US-0765739.

XX PA (IDEX-) IDEXX LAB INC.

XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
 XX DR WPI; 2002-599730/64.

XX PS New composition of matter comprising a polypeptide, useful in detecting
 XX PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
 XX PT detecting or quantifying the presence of Ehrlichia infection in mammals

XX PS Claim 1; Page 5; 29pp; English.

XX CC The invention relates to a composition of matter comprising a polypeptide
 XX CC isolated from Ehrlichia species. The composition can be used for
 XX CC detecting the presence of antibodies to Ehrlichia, comprising contacting
 XX CC one or more polypeptides with a test sample suspected of comprising
 XX CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 XX CC complexes to form and detecting the complexes, where the detection of
 XX CC polypeptide/antibody complexes is an indication that antibodies to
 XX CC Ehrlichia are present in the test sample. The composition is useful for
 XX CC detecting or quantifying the presence of E. canis or E. chaffeensis
 XX CC in mammals. The polypeptides can be used to develop monoclonal
 XX CC and/or polyclonal antibodies that can be employed in assay systems and in
 XX CC the generation of chimeric antibodies for therapeutic use or other
 XX CC similar applications. This sequence represents an E. chaffeensis peptide
 XX CC fragment used in the composition of the invention.

SQ Sequence 19 AA;
 Query Match 100.0%; Score 105; DB 23; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFIQDWDRCVIS 19
 DB 1 NTTGVGFIQDWDRCVIS 19

RESULT 2
 AAWS1092
 ID AAWS1092 standard; Protein; 286 AA.

XX AC AAWS1092;

XX DT 14-SEP-1998 (first entry)

XX DE Ehrlichia chaffeensis VSA2 protein.

XX KW MAP1 homologue; variable surface antigen; VSA2; rickettsia;
 XX KW DNA vaccine.

XX OS Ehrlichia chaffeensis.

XX FH Key Location/Qualifiers
 XX FT Peptide 1..25
 XX FT /note= "putative signal peptide"

XX PN WO9816554-A1.

XX PD 23-APR-1998.

XX PF 17-OCT-1997; 97WO-US19044.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 XX PI Nyika A, Rurangirwa FR;

XX DR WPI; 1998-251232/22.

XX DR N-PSDB; AAV07179.

XX CC Composition containing nucleic acid encoding rickettsial antigen -
 XX CC useful for, e.g. stimulating protective immune response in humans or
 XX CC animals

XX PS Claim 3; Fig 2A; 39pp; English.

XX CC This is the full-length variable surface antigen VSA2 protein of
 XX CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 XX CC partial open reading frame (ORF2) of a genomic locus (see AAV07179)
 XX CC of E. chaffeensis that was obtained on the basis of homology to the
 XX CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 XX CC This genomic locus included 5 ORFs encoding similar, but
 XX CC non-identical proteins (see AAW51091-95). A claimed composition
 XX CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 XX CC (see AAW51088-99) that elicits a protective immune response against a
 XX CC rickettsial pathogen. The nucleic acid is used, in human or
 XX CC veterinary medicine, in vaccines to protect against Rickettsia,
 XX CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 XX CC polypeptides can be used diagnostically to detect antibodies
 XX CC associated with Ehrlichia infection (claimed).

SQ Sequence 286 AA;

Query Match 100.0%; Score 105; DB 19; Length 286;
 Best Local Similarity 100.0%; Pred. No. 3.7e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
 Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 3

AA06946
 ID AAY06946 standard; Protein; 286 AA.

XX AC AAY06946;

XX DT 05-JUL-1999 (first entry)

XX DE E. chaffeensis OMP-1D protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

XX OS Ehrlichia chaffeensis.

XX PN W09913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX DR WPI; 1999-254290/21.

XX DR N-PSDB; AAX34746.

XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis

XX PS Claim 14; Fig 6B; 55pp; English.

XX CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX SQ Sequence 286 AA;

Query Match 100.0%; Score 105; DB 20; Length 286;

Best Local Similarity 100.0%; Pred. No. 3.7e-08;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19

Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 4

AAB36186

ID AAB36186 standard; Protein; 286 AA.

XX AC AAB36186;

XX DT 02-MAR-2001 (first entry)

XX DE Ehrlichia chaffeensis partial VSA2.

XX KW Ehrlichia chaffeensis; VSA2; variable surface antigen 2; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.

OS Ehrlichia chaffeensis.

XX PN W0200065063-A2.

XX PD 02-NOV-2000.

XX PF 21-APR-2000; 2000WO-US10886.

XX PR 22-APR-1999; 99US-0130725.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;

XX DR WPI; 2000-679675/66.

XX DR N-PSDB; AAC68703.

XX PT New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX PS Claim 3; Page 43-44; 63pp; English.

XX CC The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.

XX SQ Sequence 286 AA;

Query Match 100.0%; Score 105; DB 21; Length 286;

Best Local Similarity 100.0%; Pred. No. 3.7e-08;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19

Db 60 NTTGVFGIEQDWDRCVIS 78

RESULT 5

AAU04196

ID AAU04196 standard; Protein; 286 AA.

XX AC AAU04196;

XX DT 23-OCT-2001 (first entry)

XX DE Variable surface antigen 2 (VSA2) from Ehrlichia chaffeensis.

XX KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.

XX OS Ehrlichia chaffeensis.

XX PN US6251872-B1.

XX PD 26-JUN-2001.

XX PF 17-OCT-1997; 97US-0953326.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX WPI; 2001-424487/45.
 DR N-PSDB; AAS07578.
 XX
 XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures
 XX Example 3; Fig 2A-2B; 30pp; English.
 PS
 XX The sequence represents the amino acid sequence of variable surface
 CC antigen 2 (VSA2) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria rumantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.
 XX
 XX Sequence 286 AA;
 SQ
 Query Match 100.0%; Score 105; DB 22; Length 286;
 Best Local Similarity 100.0%; Pred. No. 3.7e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NTTGVGFGIEQDWDRCVIS 19
 DB 60 NTTGVGFGIEQDWDRCVIS 78
 ||||||||||||||||
 RESULT 6
 ABG77938
 ID ABG77938 standard; Protein; 286 AA.
 AC ABG77938;
 XX
 XX 15-NOV-2002 (first entry)
 DT
 DE Ehrlichia chaffeensis outer membrane protein (OMP) #4.
 XX
 XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
 KW
 XX Ehrlichia chaffeensis.
 OS
 XX US2002120115-A1.
 PN
 XX 29-AUG-2002.
 PD
 XX 28-JAN-2002; 2002US-0059964.
 PF
 XX 19-MAY-1999; 99US-0314701.
 PR
 XX (RIKI/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.
 PA
 XX Rikihisa Y, Ohashi N;
 PI
 XX WPI; 2002-618954/66.
 DR
 DR N-PSDB; ABS63279.
 XX
 XX Isolated polynucleotide encoding an outer membrane protein of E.canis
 PT or E.chaffeensis used in the diagnosis of infection
 XX Disclosure; Fig 6B; 49pp; English.
 PS

XX The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.
 XX
 XX Sequence 286 AA;
 SQ
 Query Match 100.0%; Score 105; DB 23; Length 286;
 Best Local Similarity 100.0%; Pred. No. 3.7e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NTTGVGFGIEQDWDRCVIS 19
 DB 60 NTTGVGFGIEQDWDRCVIS 78
 ||||||||||||||||
 RESULT 7
 AAU96108
 ID AAU96108 standard; Protein; 286 AA.
 XX
 XX AAU96108;
 XX
 XX 02-JUL-2002 (first entry)
 DT
 XX Ehrlichia chaffeensis OMP-1D.
 DE
 XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
 KW antibacterial.
 KW
 XX Ehrlichia chaffeensis.
 OS
 XX WO200222782-A2.
 PN
 XX 21-MAR-2002.
 PD
 XX 12-SEP-2001; 2001WO-US28759.
 PF
 XX 12-SEP-2000; 2000US-0660587.
 PR
 XX (RERE-) RES DEV FOUND.
 PA
 XX Walker DH, Yu X, McBride JW;
 PI
 XX WPI; 2002-351882/38.
 DR
 XX New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections
 XX
 XX Example 3; Figure 3; 106pp; English.
 PS
 XX The invention relates to a recombinant homologous 28 kDa immunodominant
 CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX
 XX Sequence 286 AA;
 SQ
 Query Match 100.0%; Score 105; DB 23; Length 286;
 Best Local Similarity 100.0%; Pred. No. 3.7e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NTTGVGFGIEQDWDRCVIS 19
 ||||||||||||||||

```
Db      60 NTTGVFGIEQDWDRCVIS 78
RESULT 8
AAU73415
ID      AAU73415 standard; Protein; 286 AA.
XX
XX
XX      AAU73415;
XX
XX      12-MAR-2002 (first entry)
XX
XX      Ehrlichia chaffeensis outer membrane protein P28-16.
DE
KW      Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX
XX      Ehrlichia chaffeensis.
OS
XX
XX      WO200183699-A2.
PN
XX
XX      08-NOV-2001.
XX
XX      01-MAY-2001; 2001WO-US13997.
XX
XX      01-MAY-2000; 2000US-201035P.
XX
XX      (RERE-) RES DEV FOUND.
XX
XX      Walker DH, Yu X;
PI
XX
XX      WPI; 2002-066527/09.
DR
XX
XX      Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT      P28 useful as a vaccine against Ehrlichia chaffeensis
XX
XX      Disclosure; Figure 2; 97pp; English.
XX
XX      The invention relates to isolated and purified 28-kDa outer membrane
CC      proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC      are encoded by a 28kDa outer membrane protein multigene family. P28
CC      proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
CC      is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC      Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
XX      Sequence 286 AA;
Query Match      100.0%; Score 105; DB 23; Length 286;
Best Local Similarity 100.0%; Pred. No. 3.7e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 NTTGVFGIEQDWDRCVIS 19
      |||||:|||||
Db      60 NTTGVFGIEQDWDRCVIS 78
RESULT 9
AAW51094
ID      AAW51094 standard; Protein; 280 AA.
XX
XX      AAW51094;
AC
XX
XX      14-SEP-1998 (first entry)
DT
XX
XX      Ehrlichia chaffeensis VSA4 protein.
DE
XX
XX      MAP1 homologue; variable surface antigen; VSA4; rickettsia;
KW      DNA vaccine.
PN
XX
XX      Ehrlichia chaffeensis.
OS
XX
XX      Key Location/Qualifiers
FH      Peptide 1..25
FT      /note= "putative signal peptide"
XX
```

```
PN      WO9816554-A1.
XX
XX      23-APR-1998.
XX
XX      17-OCT-1997; 97WO-US19044.
PF
XX
XX      17-OCT-1996; 96US-0733230.
PR
XX
XX      (UYFL ) UNIV FLORIDA.
PA
XX
XX      Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
PI      Nyika A, Rurangirwa FR;
XX
XX      WPI; 1998-251232/22.
DR      N-PSDB; AAV07179.
XX
XX      Composition containing nucleic acid encoding rickettsial antigen -
PT      useful for, e.g. stimulating protective immune response in humans or
PT      animals
XX
XX      Claim 3; Fig 2B; 39pp; English.
PS
XX
XX      This is the full-length variable surface antigen VSA4 protein of
CC      Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
CC      partial open reading frame (ORF4) of a genomic locus (see AAV07179)
CC      of E. chaffeensis that was obtained on the basis of homology to the
CC      major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
CC      This genomic locus included 5 ORFs encoding similar, but
CC      non-identical proteins (see AAW51091-95). A claimed composition
CC      comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
CC      (see AAW51088-99) that elicits a protective immune response against a
CC      rickettsial pathogen. The nucleic acid is used, in human or
CC      veterinary medicine, in vaccines to protect against Rickettsia,
CC      Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
CC      polypeptides can be used diagnostically to detect antibodies
CC      associated with Ehrlichia infection (claimed).
XX
XX      Sequence 280 AA;
Query Match      73.3%; Score 77; DB 19; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.00079;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 NTTGVFGIEQDWDRCVIS 19
      |||||:|||||
Db      60 NTTGVFGIEQDWDRCVIS 78
RESULT 10
AAB36188
ID      AAB36188 standard; Protein; 280 AA.
XX
XX      AAB36188;
AC
XX
XX      02-MAR-2001 (first entry)
DT
XX
XX      Ehrlichia chaffeensis partial VSA4.
DE
XX
XX      Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
KW      major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW      Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
KW      3gdorf3.
XX
XX      Ehrlichia chaffeensis.
OS
XX
XX      WO200065063-A2.
PN
XX
XX      02-NOV-2000.
PD
XX
XX      21-APR-2000; 2000WO-US10886.
PF
XX
XX      22-APR-1999; 99US-0130725.
PR
XX
```

PA (UYFL) UNIV FLORIDA.
 XX Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Alleman AR;
 XX WPI; 2000-679675/66.
 DR N-PSDB; AAC68705.
 XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 PT -
 XX Claim 3; Page 45-46; 63pp; English.
 PS The present sequence shows a high degree of similarity to the major
 XX antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhwrif3, 4hwrif1, 18hwrif1
 CC and 3gdrif3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.
 XX SQ Sequence 280 AA;
 Query Match 73.3%; Score 77; DB 21; Length 280;
 Best Local Similarity 68.4%; Pred. No. 0.00079;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVFGIEQDWDRCVIS 19
 III:IIII:IIII II
 DB 60 NTTIGVGLKQDWDGSTIT 78
 RESULT 11
 ID AAU04198 standard; Protein; 280 AA.
 XX AAU04198;
 XX 23-OCT-2001 (first entry)
 DE Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
 XX Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX Ehrlichia chaffeensis.
 XX US6251872-B1.
 XX 26-JUN-2001.
 XX 17-OCT-1997; 97US-0953326.
 XX 17-OCT-1996; 96US-0733230.
 XX (UYFL) UNIV FLORIDA.
 PA Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX WPI; 2001-424487/45.
 DR N-PSDB; AAS07578.
 XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures -

XX Example 3; Fig 2A-2B; 30pp; English.
 XX The sequence represents the amino acid sequence of variable surface
 CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.
 XX SQ Sequence 280 AA;
 Query Match 73.3%; Score 77; DB 22; Length 280;
 Best Local Similarity 68.4%; Pred. No. 0.00079;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVFGIEQDWDRCVIS 19
 III:IIII:IIII II
 DB 60 NTTIGVGLKQDWDGSTIT 78
 RESULT 12
 ID AAU73417 standard; Protein; 280 AA.
 XX AAU73417;
 XX 12-MAR-2002 (first entry)
 DE Ehrlichia chaffeensis outer membrane protein P28-18.
 XX Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
 XX Ehrlichia chaffeensis.
 XX WO200183699-A2.
 XX 08-NOV-2001.
 XX 01-MAY-2001; 2001WO-US13997.
 XX 01-MAY-2000; 2000US-201035P.
 XX (RERE-) RES DEV FOUND.
 XX Walker DH, Yu X;
 XX WPI; 2002-066527/09.
 XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
 PT P28 useful as a vaccine against Ehrlichia chaffeensis -
 XX Disclosure; Figure 2; 97pp; English.
 XX The invention relates to isolated and purified 28-kDa outer membrane
 CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
 CC are encoded by a 28kDa outer membrane protein multigene family. P28
 CC proteins are useful as a vaccine against E.chaffeensis, DNA encoding P28
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
 XX SQ Sequence 280 AA;
 Query Match 73.3%; Score 77; DB 23; Length 280;
 Best Local Similarity 68.4%; Pred. No. 0.00079;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
 |||:||||:|||| ||
Db 60 NTTGVFGLKQDWDGSTIS 78

RESULT 13

ABG30749
ID ABG30749 standard; Peptide; 19 AA.

XX AC ABG30749;

XX DT 21-OCT-2002 (first entry)

XX DE Ehrlichia chaffeensis peptide fragment #5.

XX KW Antibody detection; monoclonal antibody; polyclonal antibody.

XX OS Ehrlichia chaffeensis.

XX PN WO200257794-A2.

XX PD 25-JUL-2002.

XX PF 16-JAN-2002; 2002WO-US01395.

XX PR 18-JAN-2001; 2001US-0765739.

XX PA (IDEX-) IDEXX LAB INC.

XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;

XX WPI; 2002-599730/64.

XX PT New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
PT

XX Claim 1; Page 5; 29pp; English.

XX The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. chaffeensis peptide
CC fragment used in the composition of the invention.

XX Sequence 19 AA;

Query Match 70.5%; Score 74; DB 23; Length 19;

Best Local Similarity 68.4%; Pred. No. 0.00011;

Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
 |||:||||:|||| ||
Db 1 NTTGVFGLKQDWDGSTIS 19

RESULT 14

AAV06948

ID .AAV06948 standard; Protein; 280 AA.

XX AAY06948;

XX 05-JUL-1999 (first entry)
XX DE E. chaffeensis OMP-1F protein.
XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.

XX OS Ehrlichia chaffeensis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX WPI; 1999-254290/21.

XX DR N-PSDB; AAX34748.

XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and

XX Ehrlichia canis

XX PS Claim 16; Fig 8B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY08959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 280 AA;

Query Match 70.5%; Score 74; DB 20; Length 280;

Best Local Similarity 68.4%; Pred. No. 0.0023;

Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19
 |||:||||:|||| ||
Db 60 NTTGVFGLKQDWDGSTIS 78

RESULT 15

ABG77940

ID ABG77940 standard; Protein; 280 AA.

XX AC ABG77940;

XX DT 15-NOV-2002 (first entry)

XX DE Ehrlichia chaffeensis outer membrane protein (OMP) #6.

XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

XX OS Ehrlichia chaffeensis.

XX PN US2002120115-A1.

XX PD 29-AUG-2002.

XX PF 28-JAN-2002; 2002US-0059964.

XX PR 19-MAY-1999; 99US-0314701.

XX PA (RIKI/) RIKIHISA Y.

XX (OHAS/) OHASHI N.

PI Rikihisa Y, Ohashi N;
DR WPI; 2002-618954/66.
XX N-PSDB; ABS63281.
XX Isolated polynucleotide encoding an outer membrane protein of E.canis
PT or E.chaffeensis used in the diagnosis of infection -
PT Disclosure; Fig 8B; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX
XX
SQ Sequence 280 AA;
Query Match 70.5%; Score 74; DB 23; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.0023;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVLS 19
Db 60 NTTGVFGLKQWDGSGTIS 78
RESULT 16
ID AAU96110 standard; Protein; 280 AA.
XX
XX AAU96110;
AC
XX 02-JUL-2002 (first entry)
DT
XX Ehrlichia chaffeensis OMP-1F.
DE
XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
KW antibacterial.
KW
XX Ehrlichia chaffeensis.
OS
XX WO20022782-A2.
PN
XX 21-WAR-2002.
PD
XX 12-SEP-2001; 2001WO-US28759.
PF
XX 12-SEP-2000; 2000US-0660587.
PR
XX (NRE-) RES DEV FOUND.
PA
XX Walker DH, Yu X, McBride JW;
PI WPI; 2002-351882/38.
XX
XX New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
PT
XX
PS Example 3; Figure 3; 106pp; English.
XX
XX The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.

XX
SQ Sequence 280 AA;
Query Match 70.5%; Score 74; DB 23; Length 280;
Best Local Similarity 68.4%; Pred. No. 0.0023;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVLS 19
Db 60 NTTGVFGLKQWDGSGTIS 78
RESULT 17
ID ABG30745 standard; Peptide; 20 AA.
XX
XX ABG30745;
AC
XX 21-OCT-2002 (first entry)
DT
XX Ehrlichia chaffeensis peptide fragment #1.
DE
XX Antibody detection; monoclonal antibody; polyclonal antibody.
KW
XX Ehrlichia chaffeensis.
OS
XX WO200257794-A2.
PN
XX 25-JUL-2002.
PD
XX 16-JAN-2002; 2002WO-US01395.
PF
XX 18-JAN-2001; 2001US-0765739.
PR
XX (IDEX-) IDEXX LAB INC.
PA
XX Lawton R, O'Connor TP, Bartol BA, Machenry PS;
PI WPI; 2002-599730/64.
XX
XX New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
PT
XX
XX Claim 1; Page 5; 29pp; English.
XX
XX The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. chaffeensis peptide
CC fragment used in the composition of the invention.
XX
SQ Sequence 20 AA;
Query Match 69.5%; Score 73; DB 23; Length 20;
Best Local Similarity 68.4%; Pred. No. 0.00017;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVLS 19
Db 1 NTTGVFGLKQWDGSGAIS 19

RESULT 18
 AAY06942
 ID AAY06942 standard; Protein; 256 AA.
 XX
 AC AAY06942;

XX 05-JUL-1999 (first entry)
 XX
 DE E. chaffeensis p28 protein.

XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

XX Ehrlichia chaffeensis.

XX W09913720-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

XX 19-SEP-1997; 97US-0059353.

XX (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI; 1999-254290/21.

XX N-PSDB; AAX34742.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and

PT Ehrlichia canis

XX Claim 18; Fig 1; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the p30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 256 AA;

Query Match 69.5%; Score 73; DB 20; Length 256;
 Best Local Similarity 68.4%; Pred. No. 0.003;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVFGIEQDWDRCVIS 19
 |||||:::|:|:|
 Db 34 NTTGVFGKQNDGSAIS 52

RESULT 19
 ABG77966
 ID ABG77966 standard; Protein; 256 AA.
 XX
 AC ABG77966;

XX 15-NOV-2002 (first entry)

XX Protein encoded by Ehrlichia chaffeensis p28 gene.

XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection; p28; OMP-1.

XX Ehrlichia chaffeensis.

XX US2002120115-A1.

XX 29-AUG-2002.

XX 28-JAN-2002; 2002US-0059964.

XX

PR 19-MAY-1999; 99US-0314701.
 XX
 PA (RIKI/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.

XX Rikihisa Y, Ohashi N;

XX WPI; 2002-618954/66.

XX N-PSDB; ABS63307.

XX Isolated polynucleotide encoding an outer membrane protein of E. canis
 PT or E. chaffeensis used in the diagnosis of infection -

XX Disclosure; Fig 1; 49pp; English.

XX The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents the Ehrlichia chaffeensis OMP-1 protein encoded
 CC by the p28 gene.

XX Sequence 256 AA;

Query Match 69.5%; Score 73; DB 23; Length 256;

Best Local Similarity 68.4%; Pred. No. 0.003;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVFGIEQDWDRCVIS 19

|||||:::|:|:|

Db 34 NTTGVFGKQNDGSAIS 52

RESULT 20

AAW51095

ID AAW51095 standard; Protein; 276 AA.

XX AC AAW51095;

XX 14-SEP-1998 (first entry)

XX Ehrlichia chaffeensis VSA5 protein (partial sequence).

KW MAP1 homologue; variable surface antigen; VSA5; rickettsia;
 KW DNA vaccine.

OS Ehrlichia chaffeensis.

XX FH Key Location/Qualifiers

FT Peptide 1..25

FT /note= "putative signal peptide"

XX WO9816554-A1.

XX 23-APR-1998.

XX 17-OCT-1997; 97WO-US19044.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;

XX Nyika A, Rurangirwa FR;

XX WPI; 1998-251232/22.

XX N-PSDB; AAV07179.

XX Composition containing nucleic acid encoding rickettsial antigen -

PT useful for, e.g. stimulating protective immune response in humans or
 PT animals

PS Claim 3; Fig 2B; 39pp; English.

XX This is the near full-length variable surface antigen VSA5 protein
 CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
 CC residues. The VSA5 amino acid sequence was deduced from a partial
 CC open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
 CC chaffeensis that was obtained on the basis of homology to the major
 CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
 CC genomic locus included 5 ORFs encoding similar, but non-identical
 CC proteins (see AAW51091-95). A claimed composition comprises a
 CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
 CC that elicits a protective immune response against a rickettsial
 CC pathogen. The nucleic acid is used, in human or veterinary
 CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,
 CC Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).

XX Sequence 276 AA;

Query Match 69.5%; Score 73; DB 19; Length 276;
 Best Local Similarity 68.4%; Pred. No. 0.0032;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVGFGIEQDWDRCVIS 19
 |||||:::|:|:|
 Db 59 NTTGVGFGKQNWGSAIS 77

RESULT 21

AAB36189
 ID AAB36189 standard; Protein; 276 AA.

XX AAB36189;

XX 02-MAR-2001 (first entry)

XX Ehrlichia chaffeensis partial VSA5.

XX Ehrlichia chaffeensis; VSA5: variable surface antigen 5; MAP1:

KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;

KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;

KW 3gdorf3.

XX Ehrlichia chaffeensis.

XX WO200065063-A2.

XX 02-NOV-2000.

XX 21-APR-2000; 2000WO-US10886.

XX 22-APR-1999; 99US-0130725.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Bowie MV, Ganta RR, Burridge MJ, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 XX WPI: 2000-679675/66.

XX N-PSDB; AAC68706.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX Claim 3; Page 47; 63pp; English.

XX The present sequence shows a high degree of similarity to the major

CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccine to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid response
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.

XX Sequence 276 AA;

Query Match 69.5%; Score 73; DB 21; Length 276;

Best Local Similarity 68.4%; Pred. No. 0.0032;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 NTTGVGFGIEQDWDRCVIS 19
 |||||:::|:|:|
 Db 59 NTTGVGFGKQNWGSAIS 77

RESULT 22

AAU04199
 ID AAU04199 standard; Protein; 276 AA.

XX AAU04199;

XX 23-OCT-2001 (first entry)

XX Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.

XX Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;

KW infection; heartwater; diagnostic; variable surface antigen; VSA.

XX Ehrlichia chaffeensis.

XX US6251872-B1.

XX 26-JUN-2001.

XX 17-OCT-1997; 97US-0953326.

XX 17-OCT-1996; 96US-0733230.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;

XX WPI: 2001-424487/45.

XX N-PSDB; AAS07578.

XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures

XX Example 3; Fig 2A-2B; 30pp; English.

XX The sequence represents the amino acid sequence of variable surface
 CC antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsial infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is

CC needed.

SQ Sequence 276 AA;

Query Match 69.5%; Score 73; DB 22; Length 276;

Best Local Similarity 68.4%; Pred. No. 0.0032;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19

|||||||:|:| ||

DB 59 NTTGVFGLKQNDGSAIS 77

RESULT 23

AAW51089

ID AAW51089 standard; Protein; 280 AA.

XX AC AAW51089;

XX DT 14-SEP-1998 (first entry)

XX DE Ehrlichia chaffeensis major antigenic protein 1 (MAP1).

XX KW MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.

XX OS Ehrlichia chaffeensis.

XX PN WO9816554-A1.

XX PD 23-APR-1998.

XX PF 17-OCT-1997; 97WO-US19044.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;

XX PI Nyika A, Rurangirwa FR;

XX DR WPI; 1998-251232/22.

XX DR N-PSDB; AAV07177.

XX Composition containing nucleic acid encoding rickettsial antigen -
useful for, e.g. stimulating protective immune response in humans or
animals

XX Claim 3; Page 18-19; 39pp; English.

XX This polypeptide comprises the major antigen protein 1 gene (MAP1)
of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see
CC AAV07177). A claimed composition comprises a nucleic acid (see
CC AAV07177-82) encoding a polypeptide (see AAW51088-99) that elicits a
CC protective immune response against a rickettsial pathogen. The
CC nucleic acid is used, in human or veterinary medicine, in vaccines
CC to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
CC species. The nucleic acid does not replicate in the host but
CC remains episomal and capable of expressing polypeptide for at least
CC 19 mth. The Ehrlichia antigenic polypeptides can be used
CC diagnostically to detect antibodies associated with Ehrlichia
CC infection (claimed).

SQ Sequence 280 AA;

Query Match 69.5%; Score 73; DB 19; Length 280;

Best Local Similarity 68.4%; Pred. No. 0.0033;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19

|||||||:|:| ||

DB 60 NTTGVFGLKQNDGSAIS 78

RESULT 24

AAB36183

ID AAB36183 standard; Protein; 280 AA.

XX AC AAB36183;

XX DT 02-MAR-2001 (first entry)

XX DE Ehrlichia chaffeensis MAP1.

XX KW Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
KW. 4hworfl; 18hworfl; 3gdorf3.

XX OS Ehrlichia chaffeensis.

XX PN WO2000065063-A2.

XX PD 02-NOV-2000.

XX PF 21-APR-2000; 2000WO-US10886.

XX PR 22-APR-1999; 99US-0130725.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Allenman AR;

XX DR WPI; 2000-679675/66.

XX DR N-PSDB; AAC68700.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

XX Claim 3; Page 35-36; 63pp; English.

XX The present sequence is given in a specification relating to nucleic
acid vaccines containing genes to protect animals or humans against
CC rickettsial diseases caused by a organisms of Rickettsia sp., Ehrlichia
CC sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The vaccine comprises the
CC major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
CC (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
CC driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
CC ruminantium genes designated map 2, lhworf3, 4hworfl, 18hworfl and
CC 3gdorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.

SQ Sequence 280 AA;

Query Match 69.5%; Score 73; DB 21; Length 280;

Best Local Similarity 68.4%; Pred. No. 0.0033;

Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGIEQDWDRCVIS 19

|||||||:|:| ||

DB 60 NTTGVFGLKQNDGSAIS 78

RESULT 25

AAU04193

ID AAU04193 standard; Protein; 280 AA.

XX AC AAU04193;

XX DT 23-OCT-2001 (first entry)

XX DE Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.

XX Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic.
 XX Ehrlichia chaffeensis.
 OS US6251872-B1.
 PN 26-JUN-2001.
 PD 17-OCT-1997; 97US-0953326.
 PF 17-OCT-1996; 96US-0733230.
 XX (UYFL) UNIV FLORIDA.
 PA Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
 XX WPI; 2001-424487/45.
 DR N-PSDB; AAS07576.
 XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures
 XX Disclosure; Column 15-17; 30pp; English.
 XX The sequence represents the amino acid sequence of major antigenic
 CC protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
 CC polypeptides are useful as vaccines for conferring immunity to rickettsia
 CC infection, including Cowdria ruminantium causing heartwater. The MAP
 CC polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.
 XX Sequence 280 AA;
 SQ Query Match 69.5%; Score 73; DB 22; Length 280;
 Best Local Similarity 68.4%; Pred. No. 0.0033;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVFGIEQDWDRCVIS 19
 Db 60 NTTGVFGLKQNDGSAIS 78
 RESULT 26
 AAY06943
 ID AAY06943 standard; Protein; 281 AA.
 AC AAY06943;
 XX 05-JUL-1999 (first entry)
 DT E. chaffeensis OMP-1 protein.
 DE Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX Ehrlichia chaffeensis.
 OS WO9913720-A1.
 PN 25-MAR-1999.
 PD 18-SEP-1998; 98WO-US19600.
 PF

XX 19-SEP-1997; 97US-0059353.
 PR (OHIS) UNIV OHIO STATE.
 PA Ohashi N, Rikihisa Y;
 XX WPI; 1999-254290/21.
 DR N-PSDB; AAX34743.
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 XX Disclosure; Fig 3B; 55pp; English.
 XX The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAY06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAY06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.
 XX Sequence 281 AA;
 SQ Query Match 69.5%; Score 73; DB 20; Length 281;
 Best Local Similarity 68.4%; Pred. No. 0.0033;
 Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NTTGVFGIEQDWDRCVIS 19
 Db 59 NTTGVFGLKQNDGSAIS 77
 RESULT 27
 ABG77935
 ID ABG77935 standard; Protein; 281 AA.
 AC ABG77935;
 XX 15-NOV-2002 (first entry)
 DT Ehrlichia chaffeensis outer membrane protein (OMP) #1.
 DE Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
 KW Ehrlichia chaffeensis.
 OS US2002120115-A1.
 PN 29-AUG-2002.
 PD 28-JAN-2002; 2002US-0059964.
 PF 19-MAY-1999; 99US-0314701.
 PR (RIKI/) RIKIHISA Y.
 PA (OHAS/) OHASHI N.
 XX Rikihisa Y, Ohashi N;
 PI WPI; 2002-618954/66.
 DR N-PSDB; ABS63276.
 XX Isolated polynucleotide encoding an outer membrane protein of E. canis
 PT or E. chaffeensis used in the diagnosis of infection
 XX Claim 14; Fig 3B; 49pp; English.
 XX The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting

XX WPI; 2002-599730/64.
XX New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
PT
XX
XX Claim 1; Page 5; 29pp; English.
XX The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. canis peptide
CC fragment used in the composition of the invention.
XX
XX Sequence 20 AA;
SQ
Query Match 67.6%; Score 71; DB 23; Length 20;
Best Local Similarity 66.7%; Pred. No. 0.00035;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVI 18
DB 1 NTTGVFGLKQDWDGATI 18
RESULT 31
AAI06959
ID AAY06959 standard; Protein; 288 AA.
AC AAY06959;
XX
XX 05-JUL-1999 (first entry)
XX E. canis P30 protein.
XX
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
XX Ehrlichia canis.
XX
XX WO9913720-A1.
XX
XX 25-MAR-1999.
XX 18-SEP-1998; 98WO-US19600.
XX 19-SEP-1997; 97US-0059353.
XX (OHIS) UNIV OHIO STATE.
XX Ohashi N, Rikihisa Y;
XX
XX N-PSDB; AAX34759.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
XX Disclosure; Fig 19B; 55pp; English.
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 288 AA;
SQ
Query Match 67.6%; Score 71; DB 20; Length 288;
Best Local Similarity 66.7%; Pred. No. 0.0069;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVI 18
DB 60 NTTGVFGLKQDWDGATI 77
RESULT 32
ABG77950
ID ABG77950 standard; Protein; 288 AA.
XX
XX AC ABG77950;
XX
XX 15-NOV-2002 (first entry)
XX
XX Ehrlichia canis outer membrane protein (P30F) #1.
XX
XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
XX Ehrlichia canis.
XX
XX US2002120115-A1.
XX
XX 29-AUG-2002.
XX
XX 28-JAN-2002; 2002US-0059964.
XX
XX 19-MAY-1999; 99US-0314701.
XX
XX (RIKI/) RIKIHISA Y.
XX (OHAS/) OHASHI N.
XX
XX Rikihisa Y, Ohashi N;
XX
XX WPI; 2002-618954/66.
XX N-PSDB; ABS63291.
XX
XX Isolated polynucleotide encoding an outer membrane protein of E. canis
PT or E. chaffeensis used in the diagnosis of infection -
XX
XX Claim 10; Fig 19B; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX
XX Sequence 288 AA;
SQ
Query Match 67.6%; Score 71; DB 23; Length 288;
Best Local Similarity 66.7%; Pred. No. 0.0069;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVI 18
DB 60 NTTGVFGLKQDWDGATI 77

```
RESULT 33
AAY06962
ID AAY06962 standard; Protein; 280 AA.
XX
AC AAY06962;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis P30-2 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia canis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS ) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI; 1999-254290/21.
XX
DR N-PSDB; AAX34762.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
PS Disclosure; Fig 22B; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 280 AA;
Query Match 59.0%; Score 62; DB 20; Length 280;
Best Local Similarity 57.9%; Pred. No. 0.17;
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVIS 19
I:|||||: ||: ||
Db 59 NSTVGVFGLKHDWNGGTIS 77
RESULT 34
AAY71479
ID AAY71479 standard; Protein; 280 AA.
XX
AC AAY71479;
XX
DT 12-OCT-2000 (first entry)
XX
DE Ehrlichia canis immunoreactive protein Eca28SA3.
XX
KW Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;
KW p28 gene; polymorphic multiple gene family; immunoprotective antigen;
KW antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
KW tick-borne rickettsial disease; serodiagnostics.
XX
OS Ehrlichia canis.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
/label= signal_peptide
```

```
FT Protein 24..280
/label= Mature_Eca28SA3_28-kDa_protein
XX
PN WO200032745-A2.
XX
PD 08-JUN-2000.
XX
PF 24-NOV-1999; 99WO-US28075.
XX
PR 30-NOV-1998; 98US-0201458.
XX
PR 03-NOV-1999; 99US-0261358.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X, McBride JW;
XX
WPI; 2000-412298/35.
XX
DR N-PSDB; AAD01294, AAD01295.
XX
PT Ehrlichia canis antigens useful for vaccinating against canine
PT ehrlichiosis in dogs
XX
PS Claim 12; Page 68-69; 86pp; English.
XX
CC The patent relates to homologous 28-kilobalton (kDa) protein genes of
CC Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and
CC Eca28-2. These genes are members of a polymorphic multiple gene family
CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are
CC immunoreactive with anti-E. canis serum hence are important
CC immunoprotective antigens. The protein is useful for vaccinating
CC against E. canis infections such as canine ehrlichiosis in dogs.
CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
CC tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst
CC different strains of E. canis and hence useful for serodiagnostics of
CC canine ehrlichiosis. The present sequence is a E. canis
CC Eca28SA3 30-kDa protein which is post-translationally modified to a
CC mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX
SQ Sequence 280 AA;
Query Match 59.0%; Score 62; DB 21; Length 280;
Best Local Similarity 57.9%; Pred. No. 0.17;
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 NTTGVFGIEQDWDRCVIS 19
I:|||||: ||: ||
Db 59 NSTVGVFGLKHDWNGGTIS 77
RESULT 35
ABG77953
ID ABG77953 standard; Protein; 280 AA.
XX
AC ABG77953;
XX
DT 15-NOV-2002 (first entry)
XX
DE Ehrlichia canis outer membrane protein (P30F) #4.
XX
KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
OS Ehrlichia canis.
XX
PN US2002120115-A1.
XX
PD 29-AUG-2002.
XX
PF 28-JAN-2002; 2002US-0059964.
XX
PR 19-MAY-1999; 99US-0314701.
XX
PA (RIKI/) RIKIHISA Y.
PA (OHAS/) OHASHI N.
```

XX
PI Rikihisa Y, Ohashi N;
XX
DR WPI; 2002-618954/66.
DR N-PSDB; ABS63294.
XX
PT Isolated polynucleotide encoding an outer membrane protein of E.canis
PT or E.chaffeensis used in the diagnosis of infection -
XX
PS Claim 10; Fig 22B; 49pp; English.
XX
CC The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX
SQ Sequence 280 AA;

Query Match 59.0%; Score 62; DB 23; Length 280;
Best Local Similarity 57.9%; Pred. NO. 0.17;
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGIEODWDRCVIS 19
|:|||||:|:|
Db 59 NSTGVFGGLKHDWNGGTIS 77

Search completed: October 6, 2003, 07:48:00
Job time : 34.6045 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 24.4478 Seconds
(without alignments)
189.995 Million cell updates/sec

Title: US-09-765-739A-4

Perfect score: 98

Sequence: 1 NPTVALYGLKQDWMGVSA 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_23.*

1: sp.archaea.*

2: sp.bacteria.*

3: sp.fungi.*

4: sp.human.*

5: sp.invertebrate.*

6: sp.mammal.*

7: sp.mhc.*

8: sp.organelle.*

9: sp.phage.*

10: sp.plant.*

11: sp.rodent.*

12: sp.virus.*

13: sp.vertebrate.*

14: sp.unclassified.*

15: sp.rvirus.*

16: sp.bacteriap.*

17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	98	100.0	280	2	052104 ehrlichia c
2	98	100.0	280	2	Q8GGU5 ehrlichia c
3	98	100.0	280	2	Q8G8J3 ehrlichia c
4	88	89.8	278	2	052106 ehrlichia c
5	88	89.8	278	2	Q8G805 ehrlichia c
6	86	87.8	276	2	Q8G948 ehrlichia c
7	86	87.8	277	2	Q8G8W7 ehrlichia c
8	84	85.7	277	2	Q8GGU1 ehrlichia c
9	62	63.3	280	2	Q9ADV3 ehrlichia c
10	62	63.3	280	2	Q9F473 ehrlichia c
11	59	60.2	288	2	Q9ZGJ2 ehrlichia c
12	58	59.2	246	2	Q9RH35 ehrlichia c
13	58	59.2	275	2	Q93DD4 ehrlichia c
14	58	59.2	276	2	Q93DD1 ehrlichia c
15	58	59.2	276	2	Q85817 ehrlichia c
16	58	59.2	276	2	Q8GGU0 ehrlichia c

17	58	59.2	280	2	052107	052107 ehrlichia c
18	57	58.2	281	2	Q93DD2	Q93dd2 ehrlichia c
19	57	58.2	281	2	Q9ACI9	Q9aci9 ehrlichia c
20	55	56.1	272	2	Q9AMF6	Q9amf6 ehrlichia s
21	55	56.1	272	2	Q93E54	Q93e54 cowdria rum
22	55	56.1	284	2	Q9AFAL	Q9afal cowdria rum
23	55	56.1	284	2	Q46327	Q46327 cowdria rum
24	55	56.1	291	2	Q8G8P3	Q8g8p3 ehrlichia c
25	53	54.1	280	2	Q93DD3	Q93dd3 ehrlichia c
26	53	54.1	280	2	Q9ZGM9	Q9zgm9 ehrlichia c
27	53	54.1	280	2	085816	085816 ehrlichia c
28	52	53.1	278	2	Q9F472	Q9f472 ehrlichia c
29	52	53.1	278	2	Q9R8A8	Q9r8a8 ehrlichia c
30	52	53.1	278	2	Q9R8A7	Q9r8a7 ehrlichia c
31	52	53.1	278	2	Q9R3J3	Q9r3j3 ehrlichia c
32	52	53.1	278	2	Q9R8A6	Q9r8a6 ehrlichia c
33	52	53.1	278	2	Q9R8A9	Q9r8a9 ehrlichia c
34	52	53.1	278	2	Q9R8A5	Q9r8a5 ehrlichia c
35	52	53.1	307	2	Q9ZGJ1	Q9zgjl ehrlichia c
36	51.5	52.6	553	16	Q98PR4	Q98pr4 mycoplasma
37	51	52.0	265	2	Q9AF99	Q9af99 cowdria rum
38	51	52.0	269	2	Q93E62	Q93e62 cowdria rum
39	51	52.0	270	2	Q9AF98	Q9af98 cowdria rum
40	51	52.0	275	2	Q93E59	Q93e59 cowdria rum
41	51	52.0	276	2	Q93E60	Q93e60 cowdria rum
42	51	52.0	276	2	Q93E53	Q93e53 cowdria rum
43	51	52.0	277	2	Q93E65	Q93e65 cowdria rum
44	51	52.0	278	2	Q93E52	Q93e52 cowdria rum
45	51	52.0	278	2	Q93E57	Q93e57 cowdria rum
46	51	52.0	278	2	Q93E56	Q93e56 cowdria rum
47	51	52.0	281	2	Q9S6H1	Q9s6h1 cowdria rum
48	51	52.0	281	2	Q46328	Q46328 cowdria rum
49	51	52.0	281	2	Q9S6H0	Q9s6h0 cowdria rum
50	51	52.0	287	2	Q46329	Q46329 cowdria rum
51	51	52.0	287	2	Q9R425	Q9r425 cowdria rum
52	51	52.0	287	2	Q46331	Q46331 cowdria rum
53	51	52.0	290	2	Q46324	Q46324 cowdria rum
54	51	52.0	290	2	Q93E64	Q93e64 cowdria rum
55	51	52.0	290	2	Q46333	Q46333 cowdria rum
56	51	52.0	290	2	Q46332	Q46332 cowdria rum
57	51	52.0	290	2	Q46330	Q46330 cowdria rum
58	51	52.0	290	2	Q9AEU3	Q9aeu3 cowdria rum
59	50	51.0	286	2	052105	052105 ehrlichia c
60	50	51.0	287	2	Q8GGU3	Q8ggu3 ehrlichia c
61	50	51.0	291	2	Q8GGU2	Q8ggu2 ehrlichia c
62	50	51.0	291	2	Q8GGU1	Q8ggu1 ehrlichia c
63	50	51.0	330	2	Q08347	Q08347 streptomyc
64	49	50.0	276	2	Q9F475	Q9f475 ehrlichia c
65	49	50.0	280	2	Q93E55	Q93e55 cowdria rum
66	49	50.0	280	2	Q93E58	Q93e58 cowdria rum
67	48	49.0	439	16	Q9KLN9	Q9kln9 vibrio chol
68	47.5	48.5	612	2	Q9F7N4	Q9f7n4 uncultured
69	47	48.0	271	2	Q9AFAL	Q9afal cowdria rum
70	46	46.9	133	2	085360	085360 ehrlichia c
71	46	46.9	158	16	08KE63	08ke63 chlorobium
72	46	46.9	268	2	Q93E61	Q93e61 cowdria rum
73	46	46.9	268	2	Q93E63	Q93e63 cowdria rum
74	46	46.9	268	2	Q93E63	Q93e63 cowdria rum
75	46	46.9	308	5	Q9W3R8	Q9w3r8 drosophila
76	46	46.9	995	5	Q9VNC6	Q9vnc6 drosophila
77	46	46.9	1068	5	Q9VNC7	Q9vnc7 drosophila
78	45	45.9	461	16	08ZJ08	08zj08 versinia pe
79	45	45.9	478	16	08D150	08dls0 versinia pe
80	45	45.9	556	2	Q8GQ83	Q8gg83 pseudomonas
81	45	45.9	904	10	Q9FU09	Q9fun9 phaseolus v
82	44	44.9	477	16	Q9AAF2	Q9aaf2 caulobacter
83	44	44.9	541	16	08FLL3	08fll3 corynebacte
84	44	44.9	556	3	Q8J0P7	Q8j0p7 emericella
85	44	44.9	809	8	Q33339	Q33339 cryptoneutr
86	44	44.9	1649	16	09CFA2	Q9cfaf lactococcus
87	43.5	44.4	122	16	025081	025081 helicobacte
88	43.5	44.4	299	5	045707	045707 caenorhabdi
89	43	43.9	257	16	08XX70	08xx70 ralstonia s

90 43 43.9 279 2 Q8G801 Q8gq1 ehrlichia c
 91 43 43.9 317 16 Q8P3Y5 Q8p3y5 xanthomonas
 92 43 43.9 373 5 Q8IQW9 Q8iqw9 shewaniella
 93 43 43.9 407 16 Q8EG37 Q8eg37 shewaniella
 94 43 43.9 463 15 Q8X842 Q8x842 salmoneilla
 95 43 43.9 480 5 Q18856 Q18856 caenorhabdi
 96 43 43.9 520 2 Q8GH59 Q8gh59 clostridium
 97 43 43.9 918 3 Q01303 Q01303 neurospora
 98 43 43.9 1156 16 Q9A977 Q9a977 caulobacter
 99 42.5 43.4 328 16 Q9ZJW2 Q9zjw2 helicobacte
 100 42.5 43.4 1433 3 Q8FF17 Q8ff17 saccharomyc

ALIGNMENTS

RESULT 1
 O52104
 ID O52104 PRELIMINARY; PRT; 280 AA.
 AC 052104; O85357;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OMP-1C (28kDa outer membrane protein 15).
 GN OMP-1C.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ARKANSAS;
 RX MEDLINE=98084465; PubMed=9423849;
 RA Ohashi N., Zhi N., Zhang Y., Kikihisa Y.;
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 are encoded by a polymorphic multigene family.";
 RL Infect. Immun. 66:132-139(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;
 RX MEDLINE=98321180; PubMed=9647746;
 RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
 RA Alleman A.R.;
 RT "Molecular characterization of a 28 kDa surface antigen gene family of
 the tribe Ehrlichiae";
 RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;
 RX MEDLINE=22384137; PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL: U72291; AAC02937.1; -;
 DR EMBL: AF479833; AA012930.1; -;
 DR InterPro: IPR002566; SurfaceAg_msp4.
 DR Pfam: PF01617; SurfaceAg_2; 1.
 SQ SEQUENCE 280 AA; 30322 MW; BD835E491086DF01 CRC64;
 Query Match 100.0%; Score 98; DB 2; Length 280;
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWNGVSA 18
 DB 60 NPTVALYGLKQDWNGVSA 77
 RESULT 2
 O8GGU5
 ID O8GGU5 PRELIMINARY; PRT; 280 AA.
 AC O8GGU5
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE 28kDa outer membrane protein gene 15.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oscicola;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL: AF479839; AA012961.1; -;
 DR EMBL: AF479840; AA012967.1; -;
 SQ SEQUENCE 280 AA; 30721 MW; A36CBE26DACA2C0D CRC64;
 Query Match 100.0%; Score 98; DB 2; Length 280;
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWNGVSA 18
 DB 60 NPTVALYGLKQDWNGVSA 77
 RESULT 4
 O52106
 ID O52106 PRELIMINARY; PRT; 278 AA.
 AC O52106;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OMP-1E (28kDa outer membrane protein gene 17).
 GN OMP-1E.

Matches	15;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;
QY	1	NPTVALYGLKQDNGVSA	18						
Db	60	NPTVALYGLKQDNGEISS	77						
<p>RESULT 6</p> <p>Q8G948 PRELIMINARY; PRT; 276 AA.</p> <p>ID Q8G948 AC Q8G948</p> <p>DT 01-MAR-2003 (TrEMBLrel. 23, Created)</p> <p>DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)</p> <p>DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)</p> <p>DE 28kDa outer membrane protein gene 15.</p> <p>OS Ehrlichia chaffeensis.</p> <p>OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;</p> <p>OC Anaplasmataceae; Ehrlichia.</p> <p>NCBI_TaxID=945;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RC STRAIN=West Paces, Heartland, St. Vincent, and Wakulla;</p> <p>RX PubMed=12496165;</p> <p>RT Cheng C., Paddock C.D., Ganta R.R.;</p> <p>RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined</p> <p>RT by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes</p> <p>RT and Other Regions of the Genome.;"</p> <p>RL Infect. Immun. 71:187-195(2003).</p> <p>DR EMBL: AF479835; AAO12940.1; -.</p> <p>DR EMBL: AF479836; AAO12945.1; -.</p> <p>DR EMBL: AF479837; AAO12950.1; -.</p> <p>DR EMBL: AF479838; AAO12955.1; -.</p> <p>SQ SEQUENCE 276 AA; 30316 MW; 0D6F5353F9C0F17C CRC64;</p> <p>Query Match 87.8%; Score 86; DB 2; Length 276;</p> <p>Best Local Similarity 88.9%; Pred. No. 1.4e-05;</p> <p>Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p>									
QY	1	NPTVALYGLKQDNGVSA	18						
Db	60	NATVALYGLKQDNGVSA	77						
<p>RESULT 7</p> <p>Q8G8W7 PRELIMINARY; PRT; 277 AA.</p> <p>ID Q8G8W7 AC Q8G8W7</p> <p>DT 01-MAR-2003 (TrEMBLrel. 23, Created)</p> <p>DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)</p> <p>DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)</p> <p>DE 28kDa outer membrane protein gene 17.</p> <p>OS Ehrlichia chaffeensis.</p> <p>OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;</p> <p>OC Anaplasmataceae; Ehrlichia.</p> <p>NCBI_TaxID=945;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RC STRAIN=West Paces, Heartland, and St. Vincent;</p> <p>RX PubMed=12496165;</p> <p>RT Cheng C., Paddock C.D., Ganta R.R.;</p> <p>RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined</p> <p>RT by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes</p> <p>RT and Other Regions of the Genome.;"</p> <p>RL Infect. Immun. 71:187-195(2003).</p> <p>DR EMBL: AF479835; AAO12942.1; -.</p> <p>DR EMBL: AF479836; AAO12947.1; -.</p> <p>DR EMBL: AF479837; AAO12952.1; -.</p> <p>SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;</p> <p>Query Match 87.8%; Score 86; DB 2; Length 277;</p> <p>Best Local Similarity 88.9%; Pred. No. 1.4e-05;</p> <p>Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;</p>									

RT "Cloning and characterization of multigenes encoding the
RT Immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RC SEQUENCE FROM N.A.
RP STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis.";
RL Gene 254:245-252(2000).
DR EMBL: AF078553; AAC68667.1; -;
DR EMBL: AF082744; AAG14362.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 288 AA; 31590 MW; 86DCAECB88E9BF5E CRC64;

Query Match 60.2%; Score 59; DB 2; Length 288;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVS 17
| | :|:|:|:|:|:|:
Db 60 NTTTGVFLKQDWGAT 76

RESULT 12
Q9RH35 PRELIMINARY; PRT; 246 AA.
ID Q9RH35
AC Q9RH35
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (Fragment).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=St. Vincent;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
DR EMBL: AF077735; AAC31548.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 1
SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;

Query Match 59.2%; Score 58; DB 2; Length 246;
Best Local Similarity 60.0%; Pred. No. 0.35;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|:|:|:|:|:|:
Db 29 NTTAGVFLKQDWG 43

RESULT 13

Q93DD4 PRELIMINARY; PRT; 275 AA.
ID Q93DD4
AC Q93DD4
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V2;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393389; AAL12919.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 29974 MW; 2ECCF2F988B2E9D9 CRC64;

Query Match 59.2%; Score 58; DB 2; Length 275;
Best Local Similarity 60.0%; Pred. No. 0.39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|:|:|:|:|:|:
Db 59 NTTAGVFLKQDWG 73

RESULT 14
Q93DD1 PRELIMINARY; PRT; 276 AA.
ID Q93DD1
AC Q93DD1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF393393; AAL12923.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30028 MW; 2D7143AFCBFF2EBE CRC64;

Query Match 59.2%; Score 58; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|:|:|:|:|:|:
Db 59 NTTAGVFLKQDWG 73

RESULT 15
O85817 PRELIMINARY; PRT; 276 AA.
ID O85817
AC O85817
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28 kDa outer membrane protein).

OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RX Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V9;
RX Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Chattanooga, West Paces, Heartland, and St. Vincent;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077734; AAC31547.1; -;
DR EMBL; AF393395; AAL12925.1; -;
DR EMBL; AV117397; AM77032.1; -;
DR EMBL; AF479835; AAO12943.1; -;
DR EMBL; AF479836; AAO12948.1; -;
DR EMBL; AF479837; AAO12953.1; -;
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 59.2%; Score 58; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|||||:
Db 59 NTTAGVFLGKQDWG 73

RESULT 16

Q8GGUO
ID Q8GGUO PRELIMINARY; PRT; 276 AA.
AC Q8GGUO
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 19.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).

DR EMBL; AF479838; AAO12958.1; -.
SQ SEQUENCE 276 AA; 30045 MW; 8E3C9719D3C67A64 CRC64;
Query Match 59.2%; Score 58; DB 2; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWNG 15
| | :|||||:
Db 59 NTTAGVFLGKQDWG 73

RESULT 17

OS2107
ID OS2107 PRELIMINARY; PRT; 280 AA.
AC OS2107;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Major outer membrane protein OMP-1f (28kDa outer membrane protein gene
DE 18).
GN OMP-1f.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98084455; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas, and Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).

DR EMBL; AF479833; AAO12958.1; -.
SQ SEQUENCE 276 AA; 30045 MW; 8E3C9719D3C67A64 CRC64;

Query Match 59.2%; Score 58; DB 2; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.4;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15

```
Db      | | :|:|:|:|:|
60 NTTGVFGLKQDWG 74

RESULT 18
Q93DD2  PRELIMINARY; PRT; 281 AA.
ID Q93DD2;
AC Q93DD2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-WAR-2002 (TReMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V6;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393392; AAL12922.1; -
DR EMBL; AF393392; AAL12922.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 281 AA; 30388 MW; B54DBB55AC839A9A CRC64;

Query Match 58.2%; Score 57; DB 2; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.58;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|:|:|:|:|
Db 59 NTTGVFGLKQDWG 73

RESULT 19
Q9ACI9  PRELIMINARY; PRT; 281 AA.
ID Q9ACI9;
AC Q9ACI9;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-NAR-2003 (TReMBLrel. 23, Last annotation update)
DE Major outer membrane protein p28 (28 kDa outer membrane protein).
GN P28
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkansas;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V1;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 58.2%; Score 57; DB 2; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.58;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|:|:|:|:|
Db 59 NTTGVFGLKQDWG 73

RESULT 20
Q9AMF6  PRELIMINARY; PRT; 272 AA.
ID Q9AMF6;
AC Q9AMF6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Major antigenic protein MAPI (Fragment).
GN MAPI.
OS Ehrlichia sp. 'South African canine'.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=152574;
RN [1]
RP SEQUENCE FROM N.A.
RA Allsopp M.T., Allsopp B.A.;
RT "A novel Ehrlichia detected in dogs in South Africa.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325176; AAK14320.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 1
FT NON_TER 272
SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

Query Match 56.1%; Score 55; DB 2; Length 272;
Best Local Similarity 64.3%; Pred. No. 1.2;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGV 16
| | :|:|:|:|:|
Db 57 TKAVFGLKQDWGV 70

RESULT 21
Q93E54
```

```
ID Q93E54 PRELIMINARY; PRT; 272 AA.
AC Q93E54;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Major antigenic protein MAP1 (Fragment).
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pokoase;
RX MEDLINE=21539003; PubMed=11682561;
RA Allsopp M.T.E.P., Dorfling C.W., Maillard J.C., Bensaid A.,
RA Haydon D.T., van Heerden H., Allsopp B.A.;
RT "Ehrlichia ruminantium Major Antigenic Protein Gene (map1) Variants
RT Are Not Geographically Constrained and Show No Evidence of Having
RT Evolved under Positive Selection Pressure.";
RL J.Clin. Microbiol. 39:4200-4203(2001).
DR EMBL: AF368013; AAK98153.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 272 1
FT TER 272 1
SQ SEQUENCE 272 AA; 29225 MW; 43CEB95DEDD55BDE CRC64;

Query Match 56.1%; Score 55; DB 2; Length 272;
Best Local Similarity 64.3%; Pred. No. 1.2;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDNGV 16
| : : : : : : : : : :
Db 57 TRAVFGLKKDWDGV 70

RESULT 22
Q9AFAL PRELIMINARY; PRT; 284 AA.
ID Q9AFAL;
AC Q9AFAL;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Major antigenic protein I.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ball-3;
RA Bensaid A., Maillard J.-C., Chantal I., Perez J.-M., Martinez D.;
RT "Cowdria ruminantium major antigenic protein I (map1) gene variants
RT are not geographically constrained.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF355200; AAK27216.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 284 AA; 30720 MW; B0D3AEB9F9AB09C1 CRC64;

Query Match 56.1%; Score 55; DB 2; Length 284;
Best Local Similarity 64.3%; Pred. No. 1.2;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDNGV 16
| : : : : : : : : : :
Db 63 TRAVFGLKKDWDGV 76

RESULT 23
Q46327 PRELIMINARY; PRT; 284 AA.
ID Q46327
AC Q46327;
```

```
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE Major antigenic protein.
GN MAP1.
OS Cowdria ruminantium.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=779;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Senegal stock;
RA Van Vliet A.H.M.;
RL Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Senegal stock;
RX MEDLINE=94178956; PubMed=8132352;
RA van Vliet A.H., Jongejan F., van Kleef M., van der Zeijst B.A.;
RT "Molecular cloning, sequence analysis, and expression of the gene
RT encoding the immunodominant 32-kilodalton protein of Cowdria
RT ruminantium.";
RL Infect. Immun. 62:1451-1456(1994).
DR EMBL: X74250; CAA52309.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 284 AA; 30634 MW; 53228A889D28BEB8 CRC64;

Query Match 56.1%; Score 55; DB 2; Length 284;
Best Local Similarity 64.3%; Pred. No. 1.2;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDNGV 16
| : : : : : : : : : :
Db 63 TRAVFGLKKDWDGV 76

RESULT 24
Q8G8P3 PRELIMINARY; PRT; 291 AA.
ID Q8G8P3;
AC Q8G8P3;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Major antigenic protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479839; AAO12962.1; -.
DR EMBL: AF479840; AAO12968.1; -.
SQ SEQUENCE 291 AA; 32130 MW; 8220C436A1FC08E2 CRC64;

Query Match 56.1%; Score 55; DB 2; Length 291;
Best Local Similarity 57.1%; Pred. No. 1.2;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWN 14
| : : : : : : : : : :
Db 60 NTTGVGFIEQDWN 73

RESULT 25
Q93DD3
```

```

ID Q93DD3 PRELIMINARY; PRT; 280 AA.
AC Q93DD3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V5;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393393; AAL12921.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30372 MW; C7BBC8710BC167E9 CRC64;

Query Match 54.1%; Score 53; DB 2; Length 280;
Best Local Similarity 61.5%; Pred. No. 2.5;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
Db 61 TAGVFLKQDWDG 73

RESULT 26
Q9ZGM9 PRELIMINARY; PRT; 280 AA.
ID Q9ZGM9;
AC Q9ZGM9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jax;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., Walker D.H.;
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF077733; AAC31546.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Query Match 54.1%; Score 53; DB 2; Length 280;
Best Local Similarity 61.5%; Pred. No. 2.5;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
Db 61 TAGVFLKQDWDG 73

RESULT 27
Q85816 PRELIMINARY; PRT; 280 AA.
ID Q85816;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28kDa outer membrane protein gene
DE 19).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RX MEDLINE=99175287; PubMed=10074538;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
RT human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=91HE17;
RA Yu X.-J., Walker D.H.;
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V8, and V4;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
RT chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077732; AAC31545.1; -
DR EMBL; AF393394; AAL12924.1; -
DR EMBL; AF393390; AAL12920.1; -
DR EMBL; AF479839; AAO12964.1; -
DR EMBL; AF479840; AAO12970.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Query Match 54.1%; Score 53; DB 2; Length 280;
Best Local Similarity 61.5%; Pred. No. 2.5;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
Db 61 TAGVFLKQDWDG 73

RESULT 28
Q9F472 PRELIMINARY; PRT; 278 AA.
ID Q9F472;
AC Q9F472;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-7.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;

```

RX MEDLINE-99242757; PubMed=10225842;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 RN [2].

RP SEQUENCE FROM N.A.
 RC STRAIN=Jare;
 RX MEDLINE-20432107; PubMed=10974556;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "A conserved, transcriptionally active p28 multigene locus of
 RT Ehrlichia canis.";
 RL Gene 254:241-253(2000).
 DR EMBL: AF082746; AAC64550.2; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 278 AA; 30455 MW; 241CAAB4C56CA74 CRC64;

Query Match 53.1%; Score 52; DB 2; Length 278;
 Best Local Similarity 61.5%; Pred. No. 3.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
 DB 62 TVGVFGLKHDWDG 74

RESULT 29

Q9R8A8 ID Q9R8A8 PRELIMINARY; PRT; 278 AA.
 AC Q9R8A8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OklaHoma;
 RX MEDLINE-99242757; PubMed=10225842;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL: AF082746; AAC64552.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 DR NON_TER 278 278
 FT SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 53.1%; Score 52; DB 2; Length 278;
 Best Local Similarity 61.5%; Pred. No. 3.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
 DB 62 TVGVFGLKHDWDG 74

RESULT 30

Q9R8A7 ID Q9R8A7 PRELIMINARY; PRT; 278 AA.
 AC Q9R8A7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Demon;
 RX MEDLINE-99242757; PubMed=10225842;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL: AF082747; AAC64553.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 DR NON_TER 278 278
 FT SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 53.1%; Score 52; DB 2; Length 278;
 Best Local Similarity 61.5%; Pred. No. 3.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
 DB 62 TVGVFGLKHDWDG 74

RESULT 31

Q9R3J3 ID Q9R3J3 PRELIMINARY; PRT; 278 AA.
 AC Q9R3J3;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Florida;
 RX MEDLINE-99242757; PubMed=10225842;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL: AF082750; AAC64556.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 DR NON_TER 278 278
 FT SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 53.1%; Score 52; DB 2; Length 278;
 Best Local Similarity 61.5%; Pred. No. 3.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
 DB 62 TVGVFGLKHDWDG 74

RESULT 32

Q9R8A6 ID Q9R8A6 PRELIMINARY; PRT; 278 AA.
 AC Q9R8A6;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DJ;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL; AF082748; AAC64554.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 FT NON_TER 278 278
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;
 Query Match 53.1%; Score 52; DB 2; Length 278;
 Best Local Similarity 61.5%; Pred. No. 3.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWNG 15
 DB 62 TVGVFGLKHDWDG 74
 RESULT 33
 Q9R8A9 PRELIMINARY; PRT; 278 AA.
 AC Q9R8A9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 28 kda outer membrane protein (fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Louisiana;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL; AF082745; AAC64551.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 FT NON_TER 278 278
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;
 Query Match 53.1%; Score 52; DB 2; Length 278;
 Best Local Similarity 61.5%; Pred. No. 3.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWNG 15
 DB 62 TVGVFGLKHDWDG 74
 RESULT 34
 Q9R8A5 PRELIMINARY; PRT; 278 AA.
 AC Q9R8A5;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 28 kda outer membrane protein (fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.

OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fuzzy;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 RT antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL; AF082749; AAC64555.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 FT NON_TER 278 278
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;
 Query Match 53.1%; Score 52; DB 2; Length 278;
 Best Local Similarity 61.5%; Pred. No. 3.6;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWNG 15
 DB 62 TVGVFGLKHDWDG 74
 RESULT 35
 Q9ZGJ1 PRELIMINARY; PRT; 307 AA.
 ID Q9ZGJ1;
 AC Q9ZGJ1;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Major outer membrane protein P30-1.
 GN P30-1.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oklahoma;
 RX MEDLINE=98371112; PubMed=97054112;
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
 RT "Cloning and characterization of multigenes encoding the
 RT immunodominant 30-kilodalton major outer membrane proteins of
 RT Ehrlichia canis and application of the recombinant protein for
 RT serodiagnosis.";
 RL J. Clin. Microbiol. 36:2671-2680(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oklahoma;
 RX MEDLINE=21153566; PubMed=11254561;
 RA Ohashi N., Rikihisa Y., Unver A.;
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
 RT Membrane Protein Multigene Family in Ehrlichia canis and E.
 RT chaffeensis.";
 RL Infect. Immun. 69:2083-2091(2001).
 DR EMBL; AF078553; AAC68666.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 307 AA; 34108 MW; 9DA9FD63EBF8BC97 CRC64;
 Query Match 53.1%; Score 52; DB 2; Length 307;
 Best Local Similarity 61.5%; Pred. No. 4;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 3 TVALYGLKQDWNG 15
 DB 91 TVGVFGLKHDWDG 103
 Search completed: October 6, 2003, 07:52:36
 Job time : 26.4478 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:03 ; Search time 31.8358 Seconds
(without alignments)
89.744 Million cell updates/sec

Title: US-09-765-739A-4

Perfect score: 98

Sequence: 1 NPTVALYGLKQDWNGVSA 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_19Jun03.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98	100.0	18	23	ABG30746
2	98	100.0	280	20	AAU06945
3	98	100.0	280	23	ABG77937
4	98	100.0	280	23	AAU96107
5	98	100.0	280	23	AAU73414
6	88	89.8	18	23	ABG30748
7	88	89.8	278	19	AAW51093
8	88	89.8	278	20	AAU06947
9	88	89.8	278	21	AAAB36187

1	98				

83 S. pneumoniae type
84 Pinus radiata tran
85 C883P predicted am
86 CDX2. Homo sapien
87 Human colon tumour
88 Propionibacterium
89 Novel human diagno
90 Bacillus deramific
91 Bacillus deramific
92 Bacillus deramific
93 Bacillus deramific
94 Pullulanase amino
95 Bacillus deramific
96 Bacillus deramific
97 Bacillus deramific
98 Bacillus deramific
99 Bacillus deramific
100 Bacillus deramific

ALIGNMENTS

RESULT 1

ID ABG30746 standard; Peptide; 18 AA.

AC ABG30746;

DT 21-OCT-2002 (first entry)

DE Ehrlichia chaffeensis peptide fragment #2.

KW Antibody detection; monoclonal antibody; polyclonal antibody.

OS Ehrlichia chaffeensis.

PN WO200257794-A2.

PD 25-JUL-2002.

PF 16-JAN-2002; 2002WO-US01395.

PR 18-JAN-2001; 2001US-0765739.

PA (INDEX-) IDEXX LAB INC.

PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;

DR WPI: 2002-599730/64.

PT New composition of matter comprising a polypeptide, useful in detecting the presence of antibodies to Ehrlichia canis or chaffeensis, or in detecting or quantifying the presence of Ehrlichia infection in mammals

PS Claim 1; Page 5; 29pp; English.

CC The invention relates to a composition of matter comprising a polypeptide isolated from Ehrlichia species. The composition can be used for detecting the presence of antibodies to Ehrlichia, comprising contacting one or more polypeptides with a test sample suspected of comprising antibodies to Ehrlichia, under conditions that allow polypeptide/antibody complexes to form and detecting the complexes, where the detection of polypeptide/antibody complexes is an indication that antibodies to Ehrlichia are present in the test sample. The composition is useful for detecting or quantifying the presence of E. canis or E. chaffeensis infection in mammals. The polypeptides can be used to develop monoclonal and/or polyclonal antibodies that can be employed in assay systems and the generation of chimeric antibodies for therapeutic use or other similar applications. This sequence represents an E. chaffeensis peptide fragment used in the composition of the invention.

SQ Sequence 18 AA;

Query Match 100.0%; Score 98; DB 23; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.2e-10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18

DB 1 NPTVALYGLKQDWNGVSA 18

RESULT 2

ID AAY06945 standard; Protein; 280 AA.

XX AAY06945;

DT 05-JUL-1999 (first entry)

DE E. chaffeensis OMP-1C protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30; detection; dog.

OS Ehrlichia chaffeensis.

PN WO9913720-A1.

PD 25-MAR-1999.

PF 18-SEP-1998; 98WO-US19600.

PR 19-SEP-1997; 97US-0059353.

PA (OHIS) UNIV OHIO STATE.

PI Ohashi N, Rikihisa Y;

DR WPI: 1999-254290/21.

DR N-PSDB; AAX34745.

PT Novel outer membrane proteins from Ehrlichia chaffeensis and Ehrlichia canis

PS Claim 13; Fig 5B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part of the OMP family and consist of proteins OMP-1, -1(B to Z) shown in AAY06943-958. The E. canis proteins form part of the P30 family and consist of proteins shown in AAY06959-970. The proteins and genes are used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 280 AA;

Query Match 100.0%; Score 98; DB 20; Length 280;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18

DB 60 NPTVALYGLKQDWNGVSA 77

RESULT 3

ID ABG77937 standard; Protein; 280 AA.

XX ABG77937;

DT 15-NOV-2002 (first entry)

DE Ehrlichia chaffeensis outer membrane protein (OMP) #3.

XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX OS Ehrlichia chaffeensis.
XX PN US2002120115-A1.
XX PD 29-AUG-2002.
XX XX
XX PF 28-JAN-2002; 2002US-0059964.
XX PR 19-MAY-1999; 99US-0314701.
XX XX

XX PA (RIKI/) RIKIHISA Y.
XX PA (OHAS/) OHASHI N.
XX PI Rikihisa Y, Ohashi N;
XX XX

XX DR WPI; 2002-618954/66.
XX DR N-PSDB; ABS63278.
XX XX

XX PT Isolated polynucleotide encoding an outer membrane protein of E.canis
XX PT or E.chaffeensis used in the diagnosis of infection -
XX PS Disclosure; Fig 5B; 49pp; English.
XX XX
XX CC The invention relates to an isolated polynucleotide encoding an outer
XX CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX CC in the diagnosis of infection. An infection such as human ehrlichiosis or
XX CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX CC patient, providing a polypeptide or mixture of polypeptides, contacting
XX CC the sample with the polypeptide and assaying for the formation of a
XX CC complex between antibodies in the serum sample and the polypeptide, where
XX CC formation of a complex is indicative of infection with E. chaffeensis.
XX CC This sequence represents an Ehrlichia outer membrane protein of the
XX CC invention.
XX XX

XX SQ Sequence 280 AA;
XX XX

XX Query Match 100.0%; Score 98; DB 23; Length 280;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX

XX QY 1 NPTVALYGLKQDWNGVSA 18
XX Db |||||
XX 60 NPTVALYGLKQDWNGVSA 77

XX RESULT 4
XX AAU96107
XX ID AAU96107 standard; Protein; 280 AA.
XX AC
XX XX AAU96107;
XX XX

XX DT 02-JUL-2002 (first entry)
XX XX

XX DE Ehrlichia chaffeensis OMP-1C.
XX XX

XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX XX

XX OS Ehrlichia chaffeensis.
XX XX

XX PN WO200222782-A2.
XX XX

XX PD 21-MAR-2002.
XX XX

XX PF 12-SEP-2001; 2001WO-US28759.
XX XX

XX PR 12-SEP-2000; 2000US-0660587.
XX XX

XX PA (RERE-) RES DEV FOUND.
XX XX

PI Walker DH, Yu X, McBride JW;
XX WPI; 2002-351882/38.
XX XX

XX PT New recombinant homologous 28 kilodalton immunodominant protein from
XX PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX PS Example 3; Figure 3; 106pp; English.
XX XX

XX CC The invention relates to a recombinant homologous 28 kDa immunodominant
XX CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
XX CC dispersed in a pharmaceutically acceptable carrier, is useful for
XX CC inhibiting E. canis infection in a subject. (I) is useful in the
XX CC development of vaccines and serodiagnostics that are particularly
XX CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
XX CC represent the 28-kDa antigen amino acid sequences of the invention.
XX XX

XX SQ Sequence 280 AA;
XX XX

XX Query Match 100.0%; Score 98; DB 23; Length 280;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX

XX QY 1 NPTVALYGLKQDWNGVSA 18
XX Db |||||
XX 60 NPTVALYGLKQDWNGVSA 77

XX RESULT 5
XX AAU73414
XX ID AAU73414 standard; Protein; 280 AA.
XX XX

XX AC AAU73414;
XX XX

XX DT 12-MAR-2002 (first entry)
XX XX

XX DE Ehrlichia chaffeensis outer membrane protein P28-15.
XX KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX XX

XX OS Ehrlichia chaffeensis.
XX XX

XX PN WO200183699-A2.
XX XX

XX PD 08-NOV-2001.
XX XX

XX PF 01-MAY-2001; 2001WO-US13997.
XX XX

XX PR 01-MAY-2000; 2000US-201035P.
XX XX

XX PA (RERE-) RES DEV FOUND.
XX XX

XX PI Walker DH, Yu X;
XX XX

XX DR WPI; 2002-066527/09.
XX XX

XX PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
XX PT P28 useful as a vaccine against Ehrlichia chaffeensis -
XX XX

XX PS Disclosure; Figure 2; 97pp; English.
XX XX

XX CC The invention relates to isolated and purified 28-kDa outer membrane
XX CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
XX CC are encoded by a 28kDa outer membrane protein multigene family. P28
XX CC proteins are useful as a vaccine against E.chaffeensis. DNA encoding P28
XX CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
XX CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX XX

XX SQ Sequence 280 AA;
XX XX

XX Query Match 100.0%; Score 98; DB 23; Length 280;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX XX

QY 1 NPTVALYGLKQDNGVSA 18
 DE |||||||||||||||
 DB 60 NPTVALYGLKQDNGVSA 77

RESULT 6
 ABG30748 standard; Peptide; 18 AA.
 AC ABG30748;
 XX
 XX 21-OCT-2002 (first entry)
 DT
 XX Ehrlichia chaffeensis peptide fragment #4.
 DE
 XX Antibody detection; monoclonal antibody; polyclonal antibody.
 DE
 XX Ehrlichia chaffeensis.
 KW
 XX WO200257794-A2.
 XX
 XX 25-JUL-2002.
 PD
 XX 16-JAN-2002; 2002WO-US01395.
 XX
 XX 18-JAN-2001; 2001US-0765739.
 XX
 XX (IDEX-) IDEXX LAB INC.
 XX
 XX Lawton R, O'Connor TP, Bartol BA, Machenry PS;
 PI
 XX WPI; 2002-599730/64.
 DR
 XX
 XX New composition of matter comprising a polypeptide, useful in detecting
 PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
 PT detecting or quantifying the presence of Ehrlichia infection in mammals
 PT
 XX
 PS Claim 1; Page 5; 29pp; English.

XX The invention relates to a composition of matter comprising a polypeptide
 CC isolated from Ehrlichia species. The composition can be used for
 CC detecting the presence of antibodies to Ehrlichia, comprising contacting
 CC one or more polypeptides with a test sample suspected of comprising
 CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 CC complexes to form and detecting the complexes, where the detection of
 CC polypeptide/antibody complexes is an indication that antibodies to
 CC Ehrlichia are present in the test sample. The composition is useful for
 CC detecting or quantifying the presence of E. canis or E. chaffeensis
 CC infection in mammals. The polypeptides can be used to develop monoclonal
 CC and/or polyclonal antibodies that can be employed in assay systems and in
 CC the generation of chimeric antibodies for therapeutic use or other
 CC similar applications. This sequence represents an E. chaffeensis peptide
 CC fragment used in the composition of the invention.

QY Sequence 18 AA;
 Query Match 89.8%; Score 88; DB 23; Length 18;
 Best Local Similarity 83.3%; Pred. No. 2.8e-08;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGVSA 18
 DE |||||||||||||||
 DB 1 NPTVALYGLKQDNGVSA 18

RESULT 7
 AAW51093
 ID AAW51093 standard; Protein; 278 AA.
 XX
 XX AAW51093;
 AC
 XX

DT 14-SEP-1998 (first entry)
 XX Ehrlichia chaffeensis VSA3 protein.
 DE
 XX MAP1 homologue; variable surface antigen; VSA3; rickettsia;
 KW DNA vaccine.
 KW
 XX Ehrlichia chaffeensis.
 OS
 XX Key Location/Qualifiers
 FH 1..25
 FT Peptide /note= "putative signal peptide"
 FT
 XX WO9816554-A1.
 PN
 XX 23-APR-1998.
 PD
 XX 17-OCT-1997; 97WO-US19044.
 XX
 XX 17-OCT-1996; 96US-0733230.
 XX
 XX (UYFL) UNIV FLORIDA.
 PA
 XX Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 PI
 XX WPI; 1998-251232/22.
 DR
 XX N-PSDB; AAV07179.
 XX
 XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 PT
 XX Claim 3; Fig 2A-B; 39pp; English.
 PS
 XX This is the full-length variable surface antigen VSA3 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF5) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).

XX Sequence 278 AA;
 Query Match 89.8%; Score 88; DB 19; Length 278;
 Best Local Similarity 83.3%; Pred. No. 7.4e-07;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGVSA 18
 DE |||||||||||||||
 DB 60 NPTVALYGLKQDNGVSA 77

RESULT 8
 AAY06947
 ID AAY06947 standard; Protein; 278 AA.
 XX
 XX AAY06947;
 AC
 XX 05-JUL-1999 (first entry)
 DT
 XX E. chaffeensis OMP-1E protein.
 DE
 XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.

```

XX OS Ehrlichia chaffeensis.
XX PN WC9913720-AL.
XX PD 25-MAR-1999.
XX PF 18-SEP-1998; 98WO-US19600.
XX PR 19-SEP-1997; 97US-0059353.
XX PA (OHIS ) UNIV OHIO STATE.
XX PI Ohashi N, Rikihisa Y;
XX DR WPI; 1999-254290/21.
XX DR N-PSDB; AAX34747.
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX PT Ehrlichia canis
XX PS Claim 15; Fig 7B; 55pp; English.
XX CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX SQ Sequence 278 AA;

Query Match 89.8%; Score 88; DB 20; Length 278;
Best Local Similarity 83.3%; Pred. No. 7.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
Db ||||||||| i:i:
60 NPTVALYGLKQDWEGISS 77

RESULT 9
AAB36187
ID AAB36187 standard; Protein; 278 AA.
XX AC AAB36187;
XX DT 02-MAR-2001 (first entry)
XX DE Ehrlichia chaffeensis partial VSA3.
XX KW Ehrlichia chaffeensis; VSA3; variable surface antigen 3; MAP1;
XX KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
XX KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
XX KW 3gdorf3.
XX OS Ehrlichia chaffeensis.
XX PN WO200005063-A2.
XX PD 02-NOV-2000.
XX PF 21-APR-2000; 2000WO-US10886.
XX PR 22-APR-1999; 99US-0130725.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
XX PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX DR WPI; 2000-679675/66.
XX DR N-PSDB; AAC68704.

```

```

XX PT New polynucleotides useful as DNA vaccines for conferring immunity to
XX PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX PS Claim 3; Page 44-45; 63pp; English.
XX CC The present sequence shows a high degree of similarity to the major
XX CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
XX CC used in a vaccines to protect animals or humans against rickettsial
XX CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
XX CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
XX CC protective against the rickettsial pathogen. The nucleic acid vaccines
XX CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter
XX CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
XX CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
XX CC polypeptides are useful for detecting antibodies associated with
XX CC infection by a rickettsial pathogen whilst the polynucleotides may be
XX CC used to detect the presence of rickettsial nucleic acids.
XX SQ Sequence 278 AA;

Query Match 89.8%; Score 88; DB 21; Length 278;
Best Local Similarity 83.3%; Pred. No. 7.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
Db ||||||||| i:i:
60 NPTVALYGLKQDWEGISS 77

RESULT 10
AAU04197
ID AAU04197 standard; Protein; 278 AA.
XX AC AAU04197;
XX DT 23-OCT-2001 (first entry)
XX DE Variable surface antigen 3 (VSA3) from Ehrlichia chaffeensis.
XX KW Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
XX KW infection; heartwater; diagnostic; variable surface antigen; VSA.
XX OS Ehrlichia chaffeensis.
XX PN US6251872-B1.
XX PD 26-JUN-2001.
XX PF 17-OCT-1997; 97US-0953326.
XX PR 17-OCT-1996; 96US-0733230.
XX PA (UYFL ) UNIV FLORIDA.
XX PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
XX PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX DR WPI; 2001-424487/45.
XX DR N-PSDB; AAS07578.
XX CC New MAP2 genes and polypeptides useful as vaccines for conferring
XX CC immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX CC as molecular markers in nucleic acid analysis procedures
XX PS Example 3; Fig 2A-2B; 30pp; English.
XX CC The sequence represents the amino acid sequence of variable surface
XX CC antigen 3 (VSA3) isolated from Ehrlichia chaffeensis, which
XX CC has similarity to major antigen protein (MAP). The MAP polynucleotides
XX CC and polypeptides are useful as vaccines for conferring immunity to

```

CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC The MAP polynucleotides may be used as molecular markers in nucleic acid
CC analysis procedures, and to produce the MAP polypeptides, which may
CC be used to raise antibodies that are reactive with the polypeptides.
CC The nucleic acids may further be used as probes to identify
CC complementary sequences within other nucleic acid molecules or genomes,
CC where such probes can be applied to identify or distinguish infectious
CC strains of organisms in diagnostic procedures or in rickettsial
CC research where identification of particular organisms or strains is
CC needed.
XX Sequence 278 AA;
SQ

Query Match 89.8%; Score 88; DB 22; Length 278;
Best Local Similarity 83.3%; Pred. No. 7.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWNGVSA 18
DB 60 NPTVALYGLKQDWEGISS 77

RESULT 11
AAE31090 standard; Protein; 278 AA.
AC AAE31090;
XX
XX
XX 24-FEB-2003 (first entry)
XX Ehrlichia ruminantium protein of the invention.
XX Vaccine; immunity; rickettsial infection; spotted fever; heart water;
XX typhus; pathogen; immunostimulant; antibacterial.
XX Ehrlichia ruminantium.
XX WO200266652-A2.
XX
XX 29-AUG-2002.
XX
XX 20-FEB-2002; 2002WO-US05772.
XX
XX 20-FEB-2001; 2001US-269944P.
XX (UYFL) UNIV FLORIDA.
XX Barbet AF, Whitmore KW, Kamper SM, Simbi BH, Ganta RR,
XX Moreland AL, Mwangi DM, McGuire TC, Mahan SM;
XX WPI: 2002-723186/78.
XX New Ehrlichia ruminantium polynucleotides, useful as vaccines for
XX inducing protective immunity, and protecting animals or humans against
XX rickettsial diseases, e.g. typhus, spotted fever or heart water -
XX
XX Disclosure: Page 56-57; 206pp; English.
XX
XX The present invention relates to nucleic acid vaccines for conferring
XX immunity to rickettsial infection, including Ehrlichia ruminantium
XX (formerly Cowdria ruminantium). The invention also relates to novel
XX E. ruminantium polynucleotides and their corresponding proteins.
XX Sequences of the invention are useful for inducing immunity, particularly
XX protective immunity. They are also useful for detecting the presence of
XX E. ruminantium in a biological sample. They are useful in vaccines for
XX protecting animals or humans against rickettsial diseases, e.g. typhus,
XX spotted fever or heart water. Sequences of the invention are useful for
XX detecting antibodies to pathogens. The present sequence is E. ruminantium
XX protein of the invention.
XX
XX Sequence 278 AA;
SQ

Query Match 89.8%; Score 88; DB 23; Length 278;
Best Local Similarity 83.3%; Pred. No. 7.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWNGVSA 18
DB 60 NPTVALYGLKQDWEGISS 77

RESULT 12
ABG77939 standard; Protein; 278 AA.
ID ABG77939
XX
XX AC ABG77939;
XX
XX 15-NOV-2002 (first entry)
XX Ehrlichia chaffeensis outer membrane protein (OMP) #5.
XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX Ehrlichia chaffeensis.
XX US2002120115-A1.
XX
XX 29-AUG-2002.
XX
XX 28-JAN-2002; 2002US-0059964.
XX
XX 19-MAY-1999; 99US-0314701.
XX (RIKI/) RIKIHISA Y.
XX (OHAS/) OHASHI N.
XX Rikihisa Y, Ohashi N;
XX WPI: 2002-618954/66.
XX N-PSDB; ABS63280.
XX
XX Isolated polynucleotide encoding an outer membrane protein of E. canis
XX or E. chaffeensis used in the diagnosis of infection -
XX
XX Disclosure: Fig 7B; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX in the diagnosis of infection. An infection such as human ehrlichiosis or
XX canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX patient, providing a polypeptide or mixture of polypeptides, contacting
XX the sample with the polypeptide and assaying for the formation of a
XX complex between antibodies in the serum sample and the polypeptide, where
XX formation of a complex is indicative of infection with E. chaffeensis.
XX This sequence represents an Ehrlichia outer membrane protein of the
XX invention.
XX
XX Sequence 278 AA;
SQ

Query Match 89.8%; Score 88; DB 23; Length 278;
Best Local Similarity 83.3%; Pred. No. 7.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWNGVSA 18
DB 60 NPTVALYGLKQDWEGISS 77

RESULT 13
AAU96109 standard; Protein; 278 AA.
ID AAU96109
XX
XX AC AAU96109;
XX
XX 02-JUL-2002 (first entry)
XX

```

DE Ehrlichia chaffeensis OMP-1E.
XX
KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX antibacterial.
XX
OS Ehrlichia chaffeensis.
XX
PN WO20022782-A2.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28759.
XX
PR 12-SEP-2000; 2000US-0660587.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X, McBride JW;
XX
DR WPI; 2002-351882/38.
XX
PT New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections
XX
PS Example 3; Figure 3; 106pp; English.
XX
CC The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
SQ Sequence 278 AA;

Query Match 89.8%; Score 88; DB 23; Length 278;
Best Local Similarity 83.3%; Pred. No. 7.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVSA 18
Db ||||| 1:1:
60 NPTVALYGLKQDWEGISS 77

RESULT 15
AAU06962
ID AAY06962 standard; Protein; 280 AA.
XX
AC AAY06962;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis P30-2 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia canis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS ) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI; 1999-254290/21.
DR N-PSDB; AAX34762.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
PS Disclosure; Fig 22B; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 280 AA;

Query Match 63.3%; Score 62; DB 20; Length 280;
Best Local Similarity 66.7%; Pred. No. 0.024;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
||| :||| ||||

```

```

Db      59 NSTGVGFLKHDWNG 73

RESULT 16
AAV71479
ID      AAV71479 standard; Protein; 280 AA.
XX
XX      AAV71479;
AC
XX      12-OCT-2000 (first entry)
DT
XX      Ehrlichia canis immunoreactive protein Eca28SA3.
DE
XX      Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;
KW      p28 gene; polymorphic multiple gene family; immunoprotective antigen;
KW      antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
KW      tick-borne rickettsial disease; serodiagnosis.
XX
XX      Ehrlichia canis.
OS
XX      Key      Location/Qualifiers
FH      Peptide      1..23
FT      /label= Signal_peptide
FT      Protein      24..280
FT      /label= Mature_Eca28SA3_28-kDa_protein
XX
XX      WO200032745-A2.
XX
XX      08-JUN-2000.
XX
XX      24-NOV-1999; 99WO-US28075.
XX
XX      30-NOV-1998; 98US-0201458.
XX      03-MAR-1999; 99US-0261358.
XX
XX      (RERE-) RES DEV FOUND.
XX
XX      Walker DH, Yu X, McBride JW;
PI
XX      WPI: 2000-412298/35.
DR
XX      N-PSDB: AAD01294, AAD01295.
XX
XX      Ehrlichia canis antigens useful for vaccinating against canine
XX      ehrlichiosis in dogs -
XX
XX      Claim 12; Page 68-69; 86pp; English.
XX
XX      The patent relates to homologous 28-kilobalton (kDa) protein genes of
XX      Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and
XX      Eca28-2. These genes are members of a polymorphic multiple gene family
XX      and contained in a single locus of 5.592 kb. The 28-kDa proteins are
XX      immunoreactive with anti-E. canis serum hence are important
XX      immunoprotective antigens. The protein is useful for vaccinating
XX      against E. canis infections such as canine ehrlichiosis in dogs
XX      Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
XX      tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst
XX      different strains of E. canis and hence useful for serodiagnosis of
XX      canine ehrlichiosis. The present sequence is a E. canis
XX      Eca28SA3 30-kDa protein which is post-translationally modified to a
XX      mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX
XX      Sequence      280 AA;
XX
XX      Query Match      63.3%; Score 62; DB 21; Length 280;
XX      Best Local Similarity      66.7%; Pred. No. 0.024;
XX      Matches      10; Conservative      2; Mismatches      3; Indels      0; Gaps      0;
XX
XX      Qy      1 NPTVALYGLKQDWNG 15
XX      Db      59 NSTGVGFLKHDWNG 73
XX
XX      RESULT 17
XX      AAV96102
XX      ID      AAV96102 standard; Protein; 280 AA.
XX
XX      AC      AAV96102;
XX
XX      02-JUL-2002 (first entry)
DT
XX      Ehrlichia canis p28-6.
DE
XX      Ehrlichia canis infection; vaccine; serodiagnostic; p28;
KW      antibacterial.
XX      Ehrlichia canis.
OS
XX      WO2000222782-A2.
XX
XX

```

PD 21-MAR-2002.
 XX
 PF 12-SEP-2001; 2001WO-US28759.
 XX
 PR 12-SEP-2000; 2000US-0660587.
 XX
 PA (RERE-) RES DEV FOUND.
 XX
 PI Walker DH, Yu X, McBride JW;
 DR N-PSDB; AAK68854.
 DR WPI: 2002-351882/38.
 XX
 PT New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
 PS Claim 16; Figure 7; 106pp; English.
 XX
 CC The invention relates to a recombinant homologous 28 kDa immunodominant
 CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX
 SQ Sequence 280 AA;
 Query Match 63.3%; Score 62; DB 23; Length 280;
 Best Local Similarity 66.7%; Pred. No. 0.024;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWNG 15
 DB 59 NSTVGVLKQDWNG 73
 |||::| | | | |
 |||::| | | | |
 RESULT 19
 AAWS1094
 ID AAWS1094 standard; Protein; 280 AA.
 XX
 AC AAWS1094;
 XX
 DT 14-SEP-1998 (first entry)
 DE Ehrlichia chaffeensis VSA4 protein.
 XX
 KW MAP1 homologue; variable surface antigen; VSA4; rickettsia;
 KW DNA vaccine.
 XX
 OS Ehrlichia chaffeensis.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /note= "putative signal peptide"
 XX
 PN WO9816554-A1.
 XX
 PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 DR WPI: 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX
 PT Composition containing nucleic acid encoding rickettsial antigen -

PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 PS Claim 3; Fig 2B; 39pp; English.
 XX
 CC This is the full-length variable surface antigen VSA4 protein of
 CC Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 CC partial open reading frame (ORF4) of a genomic locus (see AAV07179)
 CC of E. chaffeensis that was obtained on the basis of homology to the
 CC major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 CC This genomic locus included 5 ORFs encoding similar, but
 CC non-identical proteins (see AAW51091-95). A claimed composition
 CC comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 CC (see AAW51088-99) that elicits a protective immune response against a
 CC rickettsial pathogen. The nucleic acid is used, in human or
 CC veterinary medicine, in vaccines to protect against Rickettsia,
 CC Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 SQ Sequence 280 AA;
 Query Match 62.2%; Score 61; DB 19; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.036;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NPTVALYGLKQDWNG 15
 DB 60 NTTIGVGLKQDWNG 74
 |||::| | | | |
 |||::| | | | |
 RESULT 20
 AAB36188
 ID AAB36188 standard; Protein; 280 AA.
 XX
 AC AAB36188;
 XX
 DT 02-MAR-2001 (first entry)
 DE Ehrlichia chaffeensis partial VSA4.
 XX
 KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
 KW 3gdorf3.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO2000065063-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 21-APR-2000; 2000WO-US10886.
 XX
 PR 22-APR-1999; 99US-0130725.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 XX
 DR WPI: 2000-679675/66.
 DR N-PSDB; AAC68705.
 XX
 PT New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 XX
 PS Claim 3; Page 45-46; 63pp; English.
 XX
 CC The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be

CC used in a vaccines to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminantium genes designated map 2, lhwerf3, 4hwerf1, 19hwerf1
 CC and 3gdorf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.

XX Sequence 280 AA;
 SQ

Query Match 62.2%; Score 61; DB 21; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.036; 2; Indels 0; Gaps 0;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 Db 60 NTTIGVGLKQDWNG 74

RESULT 21
 AAU04198
 ID AAU04198 standard; Protein; 280 AA.
 AC AAU04198;
 XX
 XX 23-OCT-2001 (first entry)
 XX
 XX Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
 DE
 DE Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 KW
 XX Ehrlichia chaffeensis.
 OS
 XX US6251872-B1.
 PN
 XX 26-JUN-2001.
 PD
 XX 17-OCT-1997; 97US-0953326.
 PF
 XX 17-OCT-1996; 96US-0733230.
 PR
 XX (UYFL) UNIV FLORIDA.
 PA
 XX Barbet AF, Ganta RR, McGuire TC, BurrIDGE MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, Bowie MV, Allemen AR;
 PI
 XX WPI; 2001-424487/45.
 DR
 DR N-PSDB; AAS07578.
 DR
 XX New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures -
 PT
 XX Example 3; Fig 2A-2B; 30pp; English.

XX The sequence represents the amino acid sequence of variable surface
 CC antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminantium causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.

XX Sequence 280 AA;
 SQ

Query Match 62.2%; Score 61; DB 22; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.036; 2; Indels 0; Gaps 0;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 Db 60 NTTIGVGLKQDWNG 74

RESULT 22
 AAU73417
 ID AAU73417 standard; Protein; 280 AA.
 XX
 XX AAU73417;
 AC
 XX 12-MAR-2002 (first entry)
 DT
 XX Ehrlichia chaffeensis outer membrane protein p28-18.
 DE
 XX Ehrlichia; outer membrane protein; p28; antibiotic; vaccine.
 KW
 XX Ehrlichia chaffeensis.
 OS
 XX WO200183699-A2.
 PN
 XX 08-NOV-2001.
 PD
 XX 01-MAY-2001; 2001WO-US13997.
 PF
 XX 01-MAY-2000; 2000US-201035P.
 PR
 XX (RERE-) RES DEV FOUND.
 PA
 XX Walker DH, Yu X;
 PI
 XX WPI; 2002-066527/09.
 DR
 XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
 PT p28 useful as a vaccine against Ehrlichia chaffeensis -
 PT
 XX Disclosure; Figure 2; 97pp; English.

XX The invention relates to isolated and purified 28-kDa outer membrane
 CC proteins (p28-1 to p28-21) of Ehrlichia chaffeensis p28 proteins
 CC are encoded by a 28kDa outer membrane protein multigene family. p28
 CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding p28
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
 CC Ehrlichia chaffeensis p28 outer membrane proteins of the invention.

XX Sequence 280 AA;
 SQ

Query Match 62.2%; Score 61; DB 23; Length 280;
 Best Local Similarity 60.0%; Pred. No. 0.036;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 Db 60 NTTIGVGLKQDWNG 74

RESULT 23
 ABG30744
 ID ABG30744 standard; Peptide; 20 AA.
 XX
 XX ABG30744;
 AC
 XX 21-OCT-2002 (first entry)
 DT
 XX Ehrlichia canis peptide fragment #2.
 DE
 XX

KW Antibody detection; monoclonal antibody; polyclonal antibody.

XX Ehrlichia canis.

XX WO200257794-A2.

XX 25-JUL-2002.

XX 16-JAN-2002; 2002WO-US01395.

XX 18-JAN-2001; 2001US-0765739.

XX (IDEX-) IDEXX LAB INC.

XX Lawton R, O'Connor TP, Bartol BA, Machenry PS;

XX WPI; 2002-599730/64.

XX New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals

PS Claim 1; Page 5; 29pp; English.

XX The invention relates to a composition of matter comprising a polypeptide
XX isolated from Ehrlichia species. The composition can be used for
XX detecting the presence of antibodies to Ehrlichia, comprising contacting
XX one or more polypeptides with a test sample suspected of comprising
XX antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
XX complexes to form and detecting the complexes, where the detection of
XX polypeptide/antibody complexes is an indication that antibodies to
XX Ehrlichia are present in the test sample. The composition is useful for
XX detecting or quantifying the presence of E. canis or E. chaffeensis
XX infection in mammals. The polypeptides can be used to develop monoclonal
XX and/or polyclonal antibodies that can be employed in assay systems and in
XX the generation of chimeric antibodies for therapeutic use or other
XX similar applications. This sequence represents an E. canis peptide
XX fragment used in the composition of the invention.

SQ Sequence 20 AA;

Query Match 60.2%; Score 59; DB 23; Length 20;
Best Local Similarity 52.9%; Pred. No. 0.0035;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVS 17

DB 1 NTTGCVGLKQDWGAT 17

RESULT 24

AA06959

ID AAY06959 standard; Protein; 288 AA.

XX AC AAY06959;

XX 05-JUL-1999 (first entry)

XX E. canis P30 protein.

KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.

XX Ehrlichia canis.

XX WO9913720-A1.

XX 25-MAR-1999.

XX 18-SEP-1998; 98WO-US19600.

XX 19-SEP-1997; 97US-0059353.

XX (OHIS) UNIV OHIO STATE.

XX Ohashi N, Rikihisa Y;

XX WPI; 1999-254290/21.

XX N-PSDB; AAX34759.

XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis

XX Disclosure; Fig 19B; 55pp; English.

XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -(B to Z) shown
XX in AAY06943-938. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 288 AA;

Query Match 60.2%; Score 59; DB 20; Length 288;
Best Local Similarity 52.9%; Pred. No. 0.083;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVS 17

DB 60 NTTGCVGLKQDWGAT 76

RESULT 25

ABG77950

ID ABG77950 standard; Protein; 288 AA.

XX AC ABG77950;

XX 15-NOV-2002 (first entry)

XX Ehrlichia canis outer membrane protein (P30F) #1.

KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

XX Ehrlichia canis.

XX US2002120115-A1.

XX 29-AUG-2002.

XX 28-JAN-2002; 2002US-0059964.

XX 19-MAY-1999; 99US-0314701.

XX (RIKI/) RIKIHISA Y.

XX (OHAS/) OHASHI N.

XX Rikihisa Y, Ohashi N;

XX WPI; 2002-618954/66.

XX N-PSDB; ABS63291.

XX Isolated polynucleotide encoding an outer membrane protein of E. canis
PT or E. chaffeensis used in the diagnosis of infection -

XX Claim 10; Fig 19B; 49pp; English.

XX The invention relates to an isolated polynucleotide encoding an outer
XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX in the diagnosis of infection. An infection such as human ehrlichiosis or
XX canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX patient, providing a polypeptide or mixture of polypeptides, contacting
XX the sample with the polypeptide and assaying for the formation of a
XX complex between antibodies in the serum sample and the polypeptide, where

XX PF 28-JAN-2002; 2002US-0059964.
XX PR 19-MAY-1999; 99US-0314701.
XX PA (RIKI/) RIKIHISA Y.
XX PA (OHAS/) OHASHI N.
XX PI Rikihisa Y, Ohashi N;
XX WPI; 2002-618954/66.
DR N-PSDB; ABS63281.
XX Isolated polynucleotide encoding an outer membrane protein of E.canis
PT or E.chaffeensis used in the diagnosis of infection -
XX Disclosure; Fig 8B; 49pp; English.
XX The invention relates to an isolated polynucleotide encoding an outer
CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
CC in the diagnosis of infection. An infection such as human ehrlichiosis or
CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX SQ Sequence 280 AA;
Query Match 59.2%; Score 58; DB 23; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.12;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWNG 15
DB 60 NTTGVFGLKQDWNG 74
RESULT 29
AAU96110
ID AAU96110 standard; Protein; 280 AA.
XX AC AAU96110;
XX DT 02-JUL-2002 (first entry)
XX DE Ehrlichia chaffeensis OMP-1F.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX OS Ehrlichia chaffeensis.
XX PN WO200222782-A2.
XX PD 21-MAR-2002.
XX PF 12-SEP-2001; 2001WO-US28759.
XX PR 12-SEP-2000; 2000US-0660587.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X, McBride JW;
XX WPI; 2002-351882/38.
XX New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX Example 3; Figure 3; 106pp; English.

XX The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX SQ Sequence 280 AA;
Query Match 59.2%; Score 58; DB 23; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.12;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 NPTVALYGLKQDWNG 15
DB 60 NTTGVFGLKQDWNG 74
RESULT 30
ABG30745
ID ABG30745 standard; Peptide; 20 AA.
XX AC ABG30745;
XX DT 21-OCT-2002 (first entry)
XX DE Ehrlichia chaffeensis peptide fragment #1.
XX KW Antibody detection; monoclonal antibody; polyclonal antibody.
XX OS Ehrlichia chaffeensis.
XX PN WO200257794-A2.
XX PD 25-JUL-2002.
XX PF 16-JAN-2002; 2002WO-US01395.
XX PR 18-JAN-2001; 2001US-0765739.
XX PA (IDEX-) IDEX LAB INC.
XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
XX WPI; 2002-599730/64.
XX New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals
XX Claim 1; Page 5; 29pp; English.
XX The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. chaffeensis peptide
CC fragment used in the composition of the invention.
XX SQ Sequence 20 AA;
Query Match 58.2%; Score 57; DB 23; Length 20;

```

Best Local Similarity 60.0%; Pred. No. 0.0077;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 1 NPTVALYGLKQDWNG 15
Db 1 NTTGVGFLKQNWGD 15

RESULT 31
AA06942
ID AAY06942 standard; Protein; 256 AA.
AC AAY06942;
XX
XX 05-JUL-1999 (first entry)
XX
XX E. chaffeensis p28 protein.
XX
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.
XX
XX Ehrlichia chaffeensis.
XX
XX WO9913720-A1.
XX
XX 25-MAR-1999.
XX
XX 18-SEP-1998; 98WO-US19600.
XX
XX 19-SEP-1997; 97US-0059353.
XX
XX (OHIS ) UNIV OHIO STATE.
XX
XX Ohashi N, Rikihisa Y;
XX
XX WPI; 1999-254290/21.
XX
XX N-PSDB; AAX34742.
XX
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX
XX Claim 18; Fig 1; 55pp; English.
XX
XX The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the p30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 256 AA;
XX
XX Query Match 58.2%; Score 57; DB 20; Length 256;
XX Best Local Similarity 60.0%; Pred. No. 0.16;
XX Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 1 NPTVALYGLKQDWNG 15
Db 34 NTTGVGFLKQNWGD 48

RESULT 32
ABG77966
ID ABG77966 standard; Protein; 256 AA.
AC ABG77966;
XX
XX 15-NOV-2002 (first entry)
XX
XX Protein encoded by Ehrlichia chaffeensis p28 gene.
XX
XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection; p28; OMP-1.
XX

```

```

OS Ehrlichia chaffeensis.
XX US2002120115-A1.
XX
XX 29-AUG-2002.
XX
XX 28-JAN-2002; 2002US-0059964.
XX
XX 19-MAY-1999; 99US-0314701.
XX
XX (RIKI/) RIKIHISA Y.
XX (OHAS/) OHASHI N.
XX
XX Rikihisa Y, Ohashi N;
XX
XX WPI; 2002-618954/66.
XX
XX N-PSDB; ABSG3307.
XX
XX Isolated polynucleotide encoding an outer membrane protein of E. canis
XX or E. chaffeensis used in the diagnosis of infection -
XX Disclosure; Fig 1; 49pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding an outer
XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX in the diagnosis of infection. An infection such as human ehrlichiosis or
XX canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX patient, providing a polypeptide or mixture of polypeptides, contacting
XX the sample with the polypeptide and assaying for the formation of a
XX complex between antibodies in the serum sample and the polypeptide, where
XX formation of a complex is indicative of infection with E. chaffeensis.
XX This sequence represents the Ehrlichia chaffeensis OMP-1 protein encoded
XX by the p28 gene.
XX
XX Sequence 256 AA;
XX
XX Query Match 58.2%; Score 57; DB 23; Length 256;
XX Best Local Similarity 60.0%; Pred. No. 0.16;
XX Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 1 NPTVALYGLKQDWNG 15
Db 34 NTTGVGFLKQNWGD 48

RESULT 33
AAW51095
ID AAW51095 standard; Protein; 276 AA.
XX
XX AAW51095;
XX
XX 14-SEP-1998 (first entry)
XX
XX Ehrlichia chaffeensis VSA5 protein (partial sequence).
XX
XX MAP1 homologue; variable surface antigen; VSA5; rickettsia;
XX DNA vaccine.
XX
XX Ehrlichia chaffeensis.
XX
XX Key Location/Qualifiers
XX Peptide 1..25
XX /note= "putative signal peptide"
XX
XX WO9816554-A1.
XX
XX 23-APR-1998.
XX
XX 17-OCT-1997; 97WO-US19044.
XX
XX 17-OCT-1996; 96US-0733230.
XX
XX (UYEL ) UNIV FLORIDA.
XX

```

```

XX PT Barbet AF, Burrigide MJ, Ganta RR, Mahan SM, McGuire TC;
PI NYika A, Rurangirwa FR;
XX WPI; 1998-251232/22.
DR N-PSDB; AAV07179.
XX
XX Composition containing nucleic acid encoding rickettsial antigen -
PT useful for, e.g. stimulating protective immune response in humans or
PT animals
XX
XX Claim 3; Fig 2B; 39pp; English.
XX
XX This is the near full-length variable surface antigen VSA5 protein
CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
CC residues. The VSA5 amino acid sequence was deduced from a partial
CC open reading frame (ORF5) of a genomic locus (see AAV07179) of E.
CC chaffeensis that was obtained on the basis of homology to the major
CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
CC genomic locus included 5 ORFs encoding similar, but non-identical
CC proteins (see AAW51091-95). A claimed composition comprises a
CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
CC that elicits a protective immune response against a rickettsial
CC pathogen. The nucleic acid is used, in human or veterinary
CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,
CC Anaplasma and Cowdria species. The Ehrlichia antigenic
CC polypeptides can be used diagnostically to detect antibodies
CC associated with Ehrlichia infection (claimed).
XX
XX Sequence 276 AA;
SQ
Query Match 58.2%; Score 57; DB 19; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.17;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
DB 59 NTTGVFGLKQNDG 73
| | | : | | | : | |
| | | : | | | : | |

RESULT 34
AAB36189
ID AAB36189 standard; Protein; 276 AA.
XX
XX AAB36189;
AC
XX
XX 02-MAR-2001 (first entry)
DT
XX
XX Ehrlichia chaffeensis partial VSA5.
DE
XX
XX Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
KW 3gdoorf3.
XX
XX Ehrlichia chaffeensis.
OS
XX
XX WO200065063-A2.
PN
XX
XX 02-NOV-2000.
PD
XX
XX 21-APR-2000; 2000WO-US10886.
PF
XX
XX 22-APR-1999; 99US-0130725.
PR
XX
XX (UYFL ) UNIV FLORIDA.
PA
XX
XX Barbet AF, Bowie MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX
XX WPI; 2000-679675/66.
DR N-PSDB; AAC68706.
XX

```

```

PT New polynucleotides useful as DNA vaccines for conferring immunity to
PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
PT
XX
XX Claim 3; Page 47; 63pp; English.
XX
XX The present sequence shows a high degree of similarity to the major
CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
CC used in a vaccine to protect animals or humans against rickettsial
CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
CC protective against the rickettsial pathogen. The nucleic acid vaccines
CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
CC Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
CC and 3gdoorf3 may be used in therapeutic and diagnostic applications. The
CC polypeptides are useful for detecting antibodies associated with
CC infection by a rickettsial pathogen whilst the polynucleotides may be
CC used to detect the presence of rickettsial nucleic acids.
XX
XX Sequence 276 AA;
SQ
Query Match 58.2%; Score 57; DB 21; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.17;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
DB 59 NTTGVFGLKQNDG 73
| | | : | | | : | |
| | | : | | | : | |

RESULT 35
AAU04199
ID AAU04199 standard; Protein; 276 AA.
XX
XX AAU04199;
AC
XX
XX 23-OCT-2001 (first entry)
DT
XX
XX Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.
DE
XX
XX Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
KW infection; heartwater; diagnostic; variable surface antigen; VSA.
XX
XX Ehrlichia chaffeensis.
OS
XX
XX US6251872-B1.
PN
XX
XX 26-JUN-2001.
PD
XX
XX 17-OCT-1997; 97US-0953326.
PF
XX
XX 17-OCT-1996; 96US-0733230.
PR
XX
XX (UYFL ) UNIV FLORIDA.
PA
XX
XX Barbet AF, Ganta RR, McGuire TC, Burrigide MJ, Nyika A;
PI Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
XX WPI; 2001-424487/45.
DR N-PSDB; AAS07578.
XX
XX New MAP2 genes and polypeptides useful as vaccines for conferring
PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
PT as molecular markers in nucleic acid analysis procedures
PT
XX
XX Example 3; Fig 2A-2B; 30pp; English.
PS
XX
XX The sequence represents the amino acid sequence of variable surface
CC antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
CC has similarity to major antigen protein (MAP). The MAP polynucleotides
CC and polypeptides are useful as vaccines for conferring immunity to
CC rickettsia infection, including Cowdria ruminantium causing heartwater.
CC

```

CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious
 CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.

XX
 SQ Sequence 276 AA;

Query Match 58.2%; Score 57; DB 22; Length 276;
 Best Local Similarity 60.0%; Pred. No. 0.17;
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
 Db 59 NTTGVFGLKQNDG 73

Search completed: October 6, 2003, 07:47:59
 Job time : 33.8358 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 10.4776 Seconds
(without alignments)
72.688 Million cell updates/sec

Title: US-09-765-739A-4

Perfect score: 98

Sequence: 1 NPTVALYGLKQDWNGVSA 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	100.0	280	4	US-09-660-587-11
2	98	100.0	280	4	US-09-261-358A-11
3	98	100.0	280	4	US-09-201-458-7
4	98	100.0	280	4	US-09-314-701-6
5	88	89.8	278	3	US-08-953-326-16
6	88	89.8	278	4	US-09-660-587-13
7	88	89.8	278	4	US-09-261-358A-13
8	88	89.8	278	4	US-09-201-458-9
9	88	89.8	278	4	US-09-314-701-10
10	62	63.3	280	4	US-09-660-587-6
11	62	63.3	280	4	US-09-261-358A-6
12	62	63.3	280	4	US-09-314-701-38
13	61	62.2	280	3	US-08-953-326-17
14	59	60.2	280	4	US-09-314-701-32
15	58	59.2	280	4	US-09-660-587-14
16	58	59.2	280	4	US-09-261-358A-14
17	58	59.2	280	4	US-09-201-458-10
18	58	59.2	280	4	US-09-314-701-12
19	57	58.2	276	3	US-08-953-326-18
20	57	58.2	280	3	US-08-733-230-4
21	57	58.2	280	3	US-08-953-326-4
22	57	58.2	281	4	US-09-660-587-9
23	57	58.2	281	4	US-09-261-358A-9
24	57	58.2	281	4	US-09-201-458-5
25	57	58.2	281	4	US-09-314-701-2
26	55	56.1	284	4	US-09-660-587-15
27	55	56.1	284	4	US-09-261-358A-15

ALIGNMENTS

RESULT 1
US-09-660-587-11
; Sequence 11, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-660-587-11

Query Match 100.0%; Score 98; DB 4; Length 280;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGVSA 18
| | | | | | | | | | | | | | | | | | | | | |
DB 60 NPTVALYGLKQDNGVSA 77

RESULT 2
US-09-261-358A-11
; Sequence 11, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-261-358A-11

Query Match 100.0%; Score 98; DB 4; Length 280;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGVSA 18
| | | | | | | | | | | | | | | | | | | | | |
DB 60 NPTVALYGLKQDNGVSA 77

RESULT 3
US-09-201-458-7

; Sequence 7, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 7
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-201-458-7

Query Match 100.0%; Score 98; DB 4; Length 280;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGVSA 18
| | | | | | | | | | | | | | | | | | | | | |
DB 60 NPTVALYGLKQDNGVSA 77

RESULT 4
US-09-314-701-6
; Sequence 6, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-6

Query Match 100.0%; Score 98; DB 4; Length 280;
Best Local Similarity 100.0%; Pred. No. 4.7e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNGVSA 18
| | | | | | | | | | | | | | | | | | | | | |
DB 60 NPTVALYGLKQDNGVSA 77

RESULT 5
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
US-08-953-326-16

; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 89.8%; Score 88; DB 3; Length 278;
Best Local Similarity 83.3%; Pred. No. 2.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQVSA 18
|||||
Db 60 NPTVALYGLKQDWEIGSS 77

RESULT 6

US-09-660-587-13
; Sequence 13, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-660-587-13

Query Match 89.8%; Score 88; DB 4; Length 278;
Best Local Similarity 83.3%; Pred. No. 2.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQVSA 18
|||||
Db 60 NPTVALYGLKQDWEIGSS 77

RESULT 7

US-09-261-358A-13
; Sequence 13, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 13

; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-261-358A-13

Query Match 89.8%; Score 88; DB 4; Length 278;
Best Local Similarity 83.3%; Pred. No. 2.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQVSA 18
|||||
Db 60 NPTVALYGLKQDWEIGSS 77

RESULT 8

US-09-201-458-9
; Sequence 9, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-201-458-9

Query Match 89.8%; Score 88; DB 4; Length 278;
Best Local Similarity 83.3%; Pred. No. 2.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQVSA 18
|||||
Db 60 NPTVALYGLKQDWEIGSS 77

RESULT 9

US-09-314-701-10
; Sequence 10, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-10

Query Match 89.8%; Score 88; DB 4; Length 278;
Best Local Similarity 83.3%; Pred. No. 2.4e-07;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNQVSA 18
|||||

```
Db 60 NPTVALYGLKQDWEGISS 77

RESULT 10
US-09-660-587-6
; Sequence 6, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-6 protein
US-09-660-587-6

Query Match 63.3%; Score 62; DB 4; Length 280;
Best Local Similarity 66.7%; Pred. No. 0.0074;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
DB 59 NSTGVGFLKHDWNG 73

RESULT 11
US-09-261-358A-6
; Sequence 6, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of ECa28SA3 protein
US-09-261-358A-6

Query Match 63.3%; Score 62; DB 4; Length 280;
Best Local Similarity 66.7%; Pred. No. 0.0074;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
DB 59 NSTGVGFLKHDWNG 73

RESULT 12
US-09-314-701-38
; Sequence 38, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 38
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-38

Query Match 63.3%; Score 62; DB 4; Length 280;
Best Local Similarity 66.7%; Pred. No. 0.0074;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
DB 59 NSTGVGFLKHDWNG 73

RESULT 13
US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match 62.2%; Score 61; DB 3; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.011;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
DB 60 NTTIGVFLKQDWNG 74

RESULT 14
US-09-314-701-32
; Sequence 32, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 38
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-32

Query Match 63.3%; Score 62; DB 4; Length 280;
Best Local Similarity 66.7%; Pred. No. 0.0074;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
DB 59 NSTGVGFLKHDWNG 73

RESULT 12
US-09-314-701-38
; Sequence 38, Application US/09314701
; Patent No. 6544517
```

; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-32

Query Match 60.2%; Score 59; DB 4; Length 288;
Best Local Similarity 52.9%; Pred. No. 0.025;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNGVS 17
| | :|||||:|
Db 60 NTTGVFGLKQDWGAT 76

RESULT 15

US-09-660-587-14
; Sequence 14, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-660-587-14

Query Match 59.2%; Score 58; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.036;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|||||:|
Db 60 NTTGVFGLKQDWG 74

RESULT 16

US-09-261-358A-14
; Sequence 14, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:

; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-261-358A-14

Query Match 59.2%; Score 58; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.036;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|||||:|
Db 60 NTTGVFGLKQDWG 74

RESULT 17

US-09-201-458-10
; Sequence 10, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 10
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-201-458-10

Query Match 59.2%; Score 58; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.036;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|||||:|
Db 60 NTTGVFGLKQDWG 74

RESULT 18

US-09-314-701-12
; Sequence 12, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-12

Query Match 59.2%; Score 58; DB 4; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.036;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
| | :|||||:|
Db 60 NTTGVFGLKQDWG 74

RESULT 19

```
US-08-953-326-18
; Sequence 18, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match      58.2%; Score 57; DB 3; Length 276;
Best Local Similarity 60.0%; Pred. No. 0.053;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNG 15
DB 59 NTTGVFGLKQNDG 73

RESULT 20
US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: 6
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whillock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
```

```
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-4

Query Match      58.2%; Score 57; DB 3; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.054;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNG 15
DB 60 NTTGVFGLKQNDG 74

RESULT 21
US-08-953-326-4
; Sequence 4, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match      58.2%; Score 57; DB 3; Length 280;
Best Local Similarity 60.0%; Pred. No. 0.054;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDNG 15
DB 60 NTTGVFGLKQNDG 74

RESULT 22
US-08-660-587-9
; Sequence 9, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
```

```
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-660-587-9

Query Match          58.2%; Score 57; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.054;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
   | | | :|:|:|:|:|
Db 59 NTTGVFGLKQNDG 73

RESULT 23
US-09-261-358A-9
; Sequence 9, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-261-358A-9

Query Match          58.2%; Score 57; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.054;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
   | | | :|:|:|:|:|
Db 59 NTTGVFGLKQNDG 73

RESULT 24
US-09-201-458-5
; Sequence 5, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-201-458-5

Query Match          58.2%; Score 57; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.054;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
   | | | :|:~|:|:|:|
Db 59 NTTGVFGLKQNDG 73

RESULT 25
US-09-314-701-2
; Sequence 2, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-2

Query Match          58.2%; Score 57; DB 4; Length 281;
Best Local Similarity 60.0%; Pred. No. 0.054;
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 NPTVALYGLKQDWNG 15
   | | | :|:~|:|:|:|
Db 59 NTTGVFGLKQNDG 73

RESULT 26
US-09-660-587-15
; Sequence 15, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-660-587-15

Query Match          56.1%; Score 55; DB 4; Length 284;
Best Local Similarity 64.3%; Pred. No. 0.12;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGV 16
   | | :|:|:|:|:|
Db 63 TKAVFGLKKDWDGV 76

RESULT 27
US-09-261-358A-15
; Sequence 15, Application US/09261358A
```

```
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-261-358A-15

Query Match
Best Local Similarity 56.1%; Score 55; DB 4; Length 284;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGV 16
Db 63 TKAVFGLKKDWDGV 76

RESULT 28
US-09-201-458-11
; Sequence 11, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-Kda Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 11
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-201-458-11

Query Match
Best Local Similarity 56.1%; Score 55; DB 4; Length 284;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGV 16
Db 63 TKAVFGLKKDWDGV 76

RESULT 29
US-09-660-587-2
; Sequence 2, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
```

```
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein
US-09-660-587-2

Query Match
Best Local Similarity 53.1%; Score 52; DB 4; Length 278;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
Db 62 TVGVFGLKHDWDG 74

RESULT 30
US-09-261-358A-2
; Sequence 2, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28-1 protein
US-09-261-358A-2

Query Match
Best Local Similarity 53.1%; Score 52; DB 4; Length 278;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
Db 62 TVGVFGLKHDWDG 74

RESULT 31
US-09-201-458-2
; Sequence 2, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kda Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/201,458A
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
```

```
; OTHER INFORMATION: amino acid sequence of a 30 kDa immunoreactive
; OTHER INFORMATION: protein of Ehrlichia canis
US-09-201-458-2

Query Match          53.1%; Score 52; DB 4; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.39;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
   ||::||| ||:|
Db 62 TVGVFGLKHDWDG 74

RESULT 32
US-09-314-701-36
; Sequence 36, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 654451710
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-36

Query Match          53.1%; Score 52; DB 4; Length 307;
Best Local Similarity 61.5%; Pred. No. 0.44;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNG 15
   ||::||| ||:|
Db 91 TVGVFGLKHDWDG 103

RESULT 33
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match          52.0%; Score 51; DB 3; Length 287;
Best Local Similarity 57.1%; Pred. No. 0.6;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGV 16
   |::|||::|||
Db 63 TQTVFGLKKDWDGV 76

RESULT 34
US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match          52.0%; Score 51; DB 3; Length 287;
Best Local Similarity 57.1%; Pred. No. 0.6;
Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVALYGLKQDWNGV 16
   |::|||::|||
Db 63 TQTVFGLKKDWDGV 76

RESULT 35
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
```

Mon Oct 6 09:43:05 2003

us-09-765-739a-4.ra

Page 10

```
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match          51.0%; Score 50; DB 3; Length 286;
Best Local Similarity 50.0%; Pred. No. 0.88;
Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 NPTVALYGLKQDWN 14
Db      60 NTIVGVFGIEQDWD 73

Search completed: October 6, 2003, 07:53:52
JOB time : 11.4776 secs
```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; Search time 27.1642 Seconds
(without alignments)
189.995 Million cell updates/sec

Title: US-09-765-739A-3

Perfect score: 107

Sequence: 1 NTTVGVFGLKQNWGSAISN 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

- 1: SPREMBL_23.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_mhc.*
- 9: sp_organelle.*
- 10: sp_plant.*
- 11: sp_prodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	107	100.0	281	Q93DD2	ehrllichia c
2	107	100.0	281	Q9ACI9	ehrllichia c
3	88	82.2	246	Q9RH35	ehrllichia c
4	88	82.2	275	Q93DD4	ehrllichia c
5	88	82.2	276	Q93DD1	ehrllichia c
6	88	82.2	276	Q93DD1	ehrllichia c
7	88	82.2	276	Q93DD1	ehrllichia c
8	88	82.2	280	Q93DD3	ehrllichia c
9	88	82.2	280	Q93DD3	ehrllichia c
10	88	82.2	280	Q93DD3	ehrllichia c
11	88	82.2	280	Q93DD3	ehrllichia c
12	82	76.6	288	Q93DD3	ehrllichia c
13	80	74.8	280	Q93DD3	ehrllichia c
14	80	74.8	280	Q93DD3	ehrllichia c
15	73	68.2	286	Q93DD3	ehrllichia c
16	73	68.2	287	Q93DD3	ehrllichia c

90 44 41.1 498 10 Q9FHH4
 91 44 41.1 1037 5 Q27766
 92 44 41.1 1649 16 Q9CFA2
 93 44 41.1 3053 12 Q8GV60
 94 44 41.1 3089 12 Q8QQA0
 95 43.5 40.7 165 16 Q9HXH6
 96 43.5 40.7 347 15 Q8QAW5
 97 43.5 40.7 360 15 Q8QAY8
 98 43 40.2 110 2 Q8VTC8
 99 43 40.2 125 2 Q8VTC3
 100 43 40.2 126 2 Q8VTC7

ALIGNMENTS

RESULT 1
 Q93DD2 ID Q93DD2 PRELIMINARY; PRT; 281 AA.
 AC Q93DD2; 2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Outer membrane protein p28.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales.
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V6;
 RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
 RT chaffeensis 28 kDa outer membrane protein multigene family.*;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF393392; AAL12922.1;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2.1;
 SQ SEQUENCE 281 AA; 30388 MW; B54DBB55AC839A9A CRC64;

Query Match 100.0%; Score 107; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNWGSAISN 20
 |||||
 Db 59 NTTGVGFLKQNWGSAISN 78

RESULT 2
 Q9ACI9 ID Q9ACI9 PRELIMINARY; PRT; 281 AA.
 AC Q9ACI9;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Major outer membrane protein p28 (28 kDa outer membrane protein).
 GN p28.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;
 RX MEDLINE=9808445; PubMed=9423849;
 RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
 RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
 are encoded by a polymorphic multigene family.*";
 RL Infect. Immun. 66:132-139(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;

RX MEDLINE=21153566; PubMed=11254561;
 RA Ohashi N., Rikihisa Y., Unver A.;
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
 Membrane Protein Multigene Family in Ehrlichia canis and E.
 chaffeensis.*";
 RL Infect. Immun. 69:2083-2091(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V1;
 RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
 RT "Allele variation and patterns of transcription of the Ehrlichia
 chaffeensis 28 kDa outer membrane protein multigene family.*";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Lithonia, Arkansas, and Oscicola;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.*";
 RL Infect. Immun. 71:187-195(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;
 RX MEDLINE=98321180; PubMed=9647746;
 RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
 RA Allemen A.R.;
 RT "Molecular characterization of a 28 kDa surface antigen gene family of
 the tribe Ehrlichiae"; Commun. 247:636-643(1998).
 RL Blochem. Biophys. Res. Commun. 247:636-643(1998).
 DR EMBL: U72291; AAK28673.1;
 DR EMBL: AF393388; AAL12918.1;
 DR EMBL: AY117396; AAM77031.1;
 DR EMBL: AF479833; AAO12932.1;
 DR EMBL: AF479834; AAO12938.1;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2.1;
 SQ SEQUENCE 281 AA; 30343 MW; A99E5F7C4459AA9A CRC64;

Query Match 100.0%; Score 107; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 6.7e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNWGSAISN 20
 |||||
 Db 59 NTTGVGFLKQNWGSAISN 78

RESULT 3
 Q9RH35 ID Q9RH35 PRELIMINARY; PRT; 246 AA.
 AC Q9RH35;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Outer membrane protein p28 (Fragment).
 GN p28.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=St. Vincent;
 RX MEDLINE=99175287; PubMed=10074538;
 RA Yu X.J., McBride J.W., Walker D.H.;
 RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
 human isolates of Ehrlichia chaffeensis.*";
 RL J. Clin. Microbiol. 37:1137-1143(1999).
 DR EMBL: AF077735; AAC31548.1;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2.1;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Arkansas;

```

SQ SEQUENCE 246 AA; 26884 MW; C9776392C5129A2F CRC64;
Query Match 82.2%; Score 88; DB 2; Length 246;
Best Local Similarity 88.9%; Pred. No. 6.2e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAI 18
Db 29 NTTAGVGLKQNDGSAI 46
||| |||||:|||||

RESULT 4
Q93DD4 PRELIMINARY; PRT; 275 AA.
AC Q93DD4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VZ;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393389; AAL12919.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 275 AA; 29974 MW; 2ECCF2F98B8B2E9D9 CRC64;

Query Match 82.2%; Score 88; DB 2; Length 275;
Best Local Similarity 88.9%; Pred. No. 7e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAI 18
Db 59 NTTAGVGLKQNDGSAI 76
||| |||||:|||||

RESULT 5
Q93DD1 PRELIMINARY; PRT; 276 AA.
AC Q93DD1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Outer membrane protein p28.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V7;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF393393; AAL12923.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30028 MW; 2D7143AFCBFF2EBE CRC64;

Query Match 82.2%; Score 88; DB 2; Length 276;
Best Local Similarity 88.9%; Pred. No. 7.1e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAI 18
Db 59 NTTAGVGLKQNDGSAI 76
||| |||||:|||||

RESULT 6
Q93DD1 PRELIMINARY; PRT; 276 AA.
AC Q93DD1;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Outer membrane protein p28 (28 kDa outer membrane protein).
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RA Yu X.-J., McBride J.W., Walker D.H.;
RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
human isolates of Ehrlichia chaffeensis.";
RL J. Clin. Microbiol. 37:1137-1143(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Sapulpa;
RA Yu X.-J., Walker D.H.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=V9;
RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
RT "Allele variation and patterns of transcription of the Ehrlichia
chaffeensis 28 kDa outer membrane protein multigene family.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Chattanooga, West Paces, Heartland, and St. Vincent;
RA PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
by Sequence Analysis of the 28-kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF077734; AAC31547.1; -
DR EMBL; AF393395; AAL12925.1; -
DR EMBL; AY117397; AAM77032.1; -
DR EMBL; AF479835; AAO12943.1; -
DR EMBL; AF479836; AAO12948.1; -
DR EMBL; AF479837; AAO12953.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 276 AA; 30027 MW; 2FD3698FCF1F60BE CRC64;

Query Match 82.2%; Score 88; DB 2; Length 276;
Best Local Similarity 88.9%; Pred. No. 7.1e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAI 18
Db 59 NTTAGVGLKQNDGSAI 76
||| |||||:|||||

RESULT 7
Q8GGU0 PRELIMINARY; PRT; 276 AA.
AC Q8GGU0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 19.
OS Ehrlichia chaffeensis.
```

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;

[1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Wakulla;
 RA PubMed=12496165;
 RX Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL: AF479838; AAL12958.1; -
 SQ SEQUENCE 276 AA; 30045 MW; 8E3C9719D3C67A64 CRC64;

Query Match 82.2%; Score 88; DB 2; Length 276;
 Best Local Similarity 88.0%; Pred. No. 7.1e-06;
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAI 18
 ||| |||||:|||||
 Db 59 NTTAGVFLKQNDGSAI 76

RESULT 8

O93DD3 PRELIMINARY; PRT; 280 AA.
 ID O93DD3
 AC O93DD3
 DT 01-DEC-2001 (TREMREL. 19, Created)
 DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
 DT 01-MAR-2002 (TREMREL. 20, Last annotation update)
 DE Outer membrane protein p28.
 OS Ehrlichia chaffeensis.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;

[1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=V5;
 RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
 RT "Allele variation and patterns of transcription of the Ehrlichia
 chaffeensis 28 kDa outer membrane protein multigene family";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF393391; AAL12921.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30372 MW; C7B8C8710BC167E9 CRC64;

Query Match 82.2%; Score 88; DB 2; Length 280;
 Best Local Similarity 80.0%; Pred. No. 7.2e-06;
 Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAISN 20
 :|| |||||:|||||
 Db 59 STTAGVFLKQNDGSAISH 78

RESULT 9

O9ZGM9 PRELIMINARY; PRT; 280 AA.
 ID O9ZGM9
 AC O9ZGM9
 DT 01-MAY-1999 (TREMREL. 10, Created)
 DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE Outer membrane protein p28.
 OS Ehrlichia chaffeensis.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Jax;
 RX MEDLINE=99175287; PubMed=10074538;

RA Yu X.-J., McBride J.W., Walker D.H.;
 RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
 human isolates of Ehrlichia chaffeensis";
 J. Clin. Microbiol. 37:1137-1143(1999).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Jax;
 RA Yu X.-J., Walker D.H.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF077733; AAC31546.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30304 MW; 91C54AC7851B77F2 CRC64;

Query Match 82.2%; Score 88; DB 2; Length 280;
 Best Local Similarity 80.0%; Pred. No. 7.2e-06;
 Matches 16; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQNDGSAISN 20
 :|| |||||:|||||
 Db 59 STTAGVFLKQNDGSAISH 78

RESULT 10

O85816 PRELIMINARY; PRT; 280 AA.
 ID O85816
 AC O85816
 DT 01-NOV-1998 (TREMREL. 08, Created)
 DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
 DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
 DE Outer membrane protein p28 (28kDa outer membrane protein gene
 DE 19).
 OS Ehrlichia chaffeensis.
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 CC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=91HE17;
 RX MEDLINE=99175287; PubMed=10074538;
 RA Yu X.-J., McBride J.W., Walker D.H.;
 RT "Genetic diversity of the 28-kilodalton outer membrane protein gene in
 human isolates of Ehrlichia chaffeensis";
 J. Clin. Microbiol. 37:1137-1143(1999).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=91HE17;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=V8, and V4;
 RA Long S.W., Zhang X.-F., Qi H., Standaert S., Walker D.H., Yu X.-J.;
 RT "Allele variation and patterns of transcription of the Ehrlichia
 chaffeensis 28 kDa outer membrane protein multigene family";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Liberty, and Jax;
 RX PubMed=12496165;

RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL: AF077732; AAC31545.1; -
 DR EMBL: AF393394; AAL12924.1; -
 DR EMBL: AF393390; AAL12920.1; -
 DR EMBL: AF479839; AAO12964.1; -
 DR EMBL: AF479840; AAO12970.1; -
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 280 AA; 30277 MW; 91C54AC78507A63F CRC64;

Db 60 NTTGVFGLKQDWDGSTIS 78

RC	S'	S
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1
21	1	1
22	1	1
23	1	1
24	1	1
25	1	1
26	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	1	1
32	1	1
33	1	1
34	1	1
35	1	1
36	1	1
37	1	1
38	1	1
39	1	1
40	1	1
41	1	1
42	1	1
43	1	1
44	1	1
45	1	1
46	1	1
47	1	1
48	1	1
49	1	1
50	1	1
51	1	1
52	1	1
53	1	1
54	1	1
55	1	1
56	1	1
57	1	1
58	1	1
59	1	1
60	1	1
61	1	1
62	1	1
63	1	1
64	1	1
65	1	1
66	1	1
67	1	1
68	1	1
69	1	1
70	1	1
71	1	1
72	1	1
73	1	1
74	1	1
75	1	1
76	1	1
77	1	1
78	1	1
79	1	1
80	1	1
81	1	1
82	1	1
83	1	1
84	1	1
85	1	1
86	1	1
87	1	1
88	1	1
89	1	1
90	1	1
91	1	1
92	1	1
93	1	1
94	1	1
95	1	1
96	1	1
97	1	1
98	1	1
99	1	1
100	1	1

CT CTCCGCUYTC TACATT T

KN [I]
RP SEQUENCE FROM N. A.

KN [I]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT Cloning and characterization of multigenes encoding the
RT immunodominant 30-kilodalton major outer membrane proteins of
RT Ehrlichia canis and application of the recombinant protein for
RT serodiagnosis."
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Oklaoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
RT Membrane Protein Multigene Family in Ehrlichia canis and E.
RT chaffeensis."
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL: AF078553; AAK28599.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30603 MW; 27238BELC7E68A91 CRC64;

Query Match 74.8%; Score 80; DB 2; Length 280;
Best Local Similarity 70.0%; Pred. No. 0.00014;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 20
DB 59 NSTGVFGLKHDWNGTISN 78
|:|||||||:|:|
|:|||||||:|:|

RESULT 14
Q9F473 ID Q9F473 PRELIMINARY; PRT; 280 AA.
AC Q9F473;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-6.
GN P28-6.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu, X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis."
RL Gene 254:245-252(2000).
DR EMBL: AF082744; AAC14361.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 280 AA; 30762 MW; BE284A4B94FE3123 CRC64;

Query Match 74.8%; Score 80; DB 2; Length 280;
Best Local Similarity 70.0%; Pred. No. 0.00014;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 20
DB 59 NSTGVFGLKHDWNGTISN 78
|:|||||||:|:|
|:|||||||:|:|

```

```

RESULT 15
Q52105 ID Q52105 PRELIMINARY; PRT; 286 AA.
AC Q52105;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OMP-1D (28kDa outer membrane protein gene 16).
GN OMP-1D.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
RT are encoded by a polymorphic multigene family."
RL Infect. Immun. 66:132-139(1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., Burridge M.J.,
RA Allemen A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
RT the tribe Ehrlichiae."
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=Osciola;
RX PubMed=12496165;
RA Cheng C., Faddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome."
RL Infect. Immun. 71:187-195(2003).
DR EMBL: U72291; AAC02938.1; -.
DR EMBL: AF479833; AAC26718.1; -.
DR EMBL: AF479834; AAO12935.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 286 AA; 31509 MW; F145A79270F386BE CRC64;

Query Match 68.2%; Score 73; DB 2; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.0018;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNDGSAIS 19
DB 60 NTTGVFGLKQNDGSAIS 78
|:|||||||:|:|
|:|||||||:|:|

RESULT 16
Q8GGU3 ID Q8GGU3 PRELIMINARY; PRT; 287 AA.
AC Q8GGU3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=St. Vincent;
RX PubMed=12496165;
RA Cheng C., Faddock C.D., Ganta R.R.;

```

RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479837; AAO12951.1; -.
SQ SEQUENCE 287 AA; 31504 MW; A9592249C83695B2 CRC64;

Query Match 68.2%; Score 73; DB 2; Length 287;
Best Local Similarity 68.4%; Pred. No. 0.0018;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||:|:|:|

RESULT 17
Q8GGU2 PRELIMINARY; PRT; 291 AA.
AC Q8GGU2;
DT 01-WAR-2003 (TReMBLrel. 23, Created)
DT 01-WAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TReMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Wakulla;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479838; AAO12956.1; -.
SQ SEQUENCE 291 AA; 32005 MW; 22D35EDED283195A CRC64;

Query Match 68.2%; Score 73; DB 2; Length 291;
Best Local Similarity 68.4%; Pred. No. 0.0018;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||:|:|:|

RESULT 18
Q8G921 PRELIMINARY; PRT; 291 AA.
AC Q8G921;
DT 01-WAR-2003 (TReMBLrel. 23, Created)
DT 01-WAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TReMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 16.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-West Paces, and Heartland;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
and Other Regions of the Genome.";
RL Infect. Immun. 71:187-195(2003).
DR EMBL: AF479835; AAO12941.1; -.
DR EMBL: AF479836; AAO12946.1; -.
SQ SEQUENCE 291 AA; 31991 MW; 22D35A2202831369 CRC64;

Query Match 68.2%; Score 73; DB 2; Length 291;
Best Local Similarity 68.4%; Pred. No. 0.0018;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAIS 19
Db 60 NTTGVFGIEQDWDRCVIS 78
|||||:|:|:|

RESULT 19
Q9F472 PRELIMINARY; PRT; 278 AA.
AC Q9F472;
DT 01-WAR-2001 (TReMBLrel. 16, Created)
DT 01-WAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE P28-7.
GN P28-7.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
RT Ehrlichia canis";
RL Gene 254:245-252(2000).
DR EMBL: AF082744; AAC64550.2; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30455 MW; 2411CAA84C56CA74 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0025;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTGVFGLKQNWGSAI 18
Db 61 STGVFGLKHDWGSPI 77
:|||||:|:|:|

RESULT 20
Q9R8A8 PRELIMINARY; PRT; 278 AA.
ID Q9R8A8;
AC Q9R8A8;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Oklahoma;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic

```
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082746; AAC64552.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0025;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTGVFGLKONWDGSAI 18
Db 61 STGVFGLKHDWDGSP 77

RESULT 21
Q9R8A7 ID Q9R8A7 PRELIMINARY; PRT; 278 AA.
AC Q9R8A7
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Demon;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082747; AAC64553.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0025;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTGVFGLKONWDGSAI 18
Db 61 STGVFGLKHDWDGSP 77

RESULT 22
Q9R3J3 ID Q9R3J3 PRELIMINARY; PRT; 278 AA.
AC Q9R3J3
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Florida;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
```

```
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082750; AAC64556.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0025;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTGVFGLKONWDGSAI 18
Db 61 STGVFGLKHDWDGSP 77

RESULT 23
Q9R8A6 ID Q9R8A6 PRELIMINARY; PRT; 278 AA.
AC Q9R8A6
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DJ;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
DR EMBL: AF082748; AAC64554.1; -.
DR InterPro: IPR002566; Surface_Ag_msp4.
DR Pfam: PF01617; Surface_Ag_2; 1.
FT NON_TER 278 278
SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.0025;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TTGVFGLKONWDGSAI 18
Db 61 STGVFGLKHDWDGSP 77

RESULT 24
Q9R8A9 ID Q9R8A9 PRELIMINARY; PRT; 278 AA.
AC Q9R8A9
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 28 kDa outer membrane protein (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Louisiana;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
RT 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
RT antigen."
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
```

DR EMBL; AF082745; AAC64551.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT NON_TER 278 278
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
 Best Local Similarity 76.5%; Pred. No. 0.0025;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 TTGVGFLKQNDGSAI 18
 :||||| :|||||
 Db 61 STVGFLKHDWDGSP 77

RESULT 25

Q9R8A5 PRELIMINARY; PRT; 278 AA.

AC Q9R8A5;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 28 kDa outer membrane protein (Fragment).
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fuzzy;
 RA MEDLINE=99242757; PubMed=10225842;
 RX McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 DR EMBL; AF082749; AAC64555.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 FT NON_TER 278 278
 SQ SEQUENCE 278 AA; 30485 MW; 697CB6CA413BBA68 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 278;
 Best Local Similarity 76.5%; Pred. No. 0.0025;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 TTGVGFLKQNDGSAI 18
 :||||| :|||||
 Db 61 STVGFLKHDWDGSP 77

RESULT 26

O9ZGJ1 PRELIMINARY; PRT; 307 AA.

AC O9ZGJ1;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Major outer membrane protein p30-1.
 GN P30-1.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OklaHoma;
 RX MEDLINE=98371112; PubMed=9705412;
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
 RT "Cloning and characterization of multigenes encoding the
 immunodominant 30-kilodalton major outer membrane proteins of
 Ehrlichia canis and application of the recombinant protein for
 serodiagnosis.";
 RL Infect. Immun. 69:2083-2091(2001).

RL J. Clin. Microbiol. 36:2671-2680(1998).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OklaHoma;
 RX MEDLINE=21153566; PubMed=11254561;
 RA Ohashi N., Rikihisa Y., Unver A.;
 RT "Analysis of Transcriptionally Active Gene Clusters of Major Outer
 Membrane Protein Multigene Family in Ehrlichia canis and E.
 chaffeensis.";
 RT Infect. Immun. 69:2083-2091(2001).
 RL EMBL; AF078553; AAC68666.1; -;
 DR InterPro: IPR002566; Surface_Ag_msp4.
 DR Pfam: PF01617; Surface_Ag_2; 1.
 SQ SEQUENCE 307 AA; 34108 MW; 9DA9FD63EBF8BC97 CRC64;

Query Match 67.3%; Score 72; DB 2; Length 307;
 Best Local Similarity 76.5%; Pred. No. 0.0028;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 TTGVGFLKQNDGSAI 18
 :||||| :|||||
 Db 90 STVGFLKHDWDGSP 106

RESULT 27

O9F475 PRELIMINARY; PRT; 276 AA.

AC O9F475;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE P28-3 (Major outer membrane protein p30-4).
 GN P28-3 OR P30-4.
 OS Ehrlichia canis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Jake;
 RX MEDLINE=99242757; PubMed=10225842;
 RA McBride J.W., Yu, Xj, Walker D.H.;
 RT "Molecular cloning of the gene for a conserved major immunoreactive
 28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
 antigen.";
 RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Jake;
 RX MEDLINE=20432107; PubMed=10974556;
 RA McBride J.W., Yu X.J., Walker D.H.;
 RT "A conserved, transcriptionally active p28 multigene locus of
 Ehrlichia canis.";
 RL Gene 254:245-252(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OklaHoma;
 RX MEDLINE=98371112; PubMed=9705412;
 RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
 RT "Cloning and characterization of multigenes encoding the
 immunodominant 30-kilodalton major outer membrane proteins of
 Ehrlichia canis and application of the recombinant protein for
 serodiagnosis.";
 RL J. Clin. Microbiol. 36:2671-2680(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OklaHoma;
 RX MEDLINE=21153566; PubMed=11254561;
 RA Ohashi N., Rikihisa Y., Unver A.;
 RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
 Membrane Protein Multigene Family in Ehrlichia canis and E.
 chaffeensis.";
 RL Infect. Immun. 69:2083-2091(2001).

DR EMBL; AF082744; AAG14358.1; -;
 DR EMBL; AF324792; AAK31313.1; -;
 DR EMBL; AF078553; AAK28697.1; -;
 DR InterPro; IPR002566; Surface_Ag_msp4.
 DR Pfam; PF01617; Surface_Ag-2; 1.
 SQ SEQUENCE 276 AA; 30659 MW; CES1AB37D17AF3A4 CRC64;

Query Match 63.6%; Score 68; DB 2; Length 276;
 Best Local Similarity 61.1%; Pred. No. 0.011;
 Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAI 18
 DB 60 NTTGIFGLKESWTGGII 77

RESULT 28
 Q8G8P3 PRELIMINARY; PRT; 291 AA.
 ID AC Q8G8P3
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DE 28kDa outer membrane protein gene 16.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Liberty, and Jax;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479839; AAO12962.1; -;
 DR EMBL; AF479840; AAO12968.1; -;
 SQ SEQUENCE 291 AA; 32130 MW; 8220C436A1FC08E2 CRC64;

Query Match 63.6%; Score 68; DB 2; Length 291;
 Best Local Similarity 63.2%; Pred. No. 0.012;
 Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAIS 19
 DB 60 NTTGVGFIQDWNRCVIS 78

RESULT 29
 Q8G948 PRELIMINARY; PRT; 276 AA.
 ID AC Q8G948
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DE 28kDa outer membrane protein gene 15.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=West Paces, Heartland, St. Vincent, and Wakulla;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479835; AAO12940.1; -;
 DR EMBL; AF479836; AAO12945.1; -;

DR EMBL; AF479837; AAO12950.1; -;
 DR EMBL; AF479838; AAO12955.1; -;
 SQ SEQUENCE 276 AA; 30316 MW; OD6F5353F9C0F17C CRC64;

Query Match 58.9%; Score 63; DB 2; Length 276;
 Best Local Similarity 52.6%; Pred. No. 0.068;
 Matches 10; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAIS 19
 DB 60 NATVALYGLKQDWNNGASAS 78

RESULT 30
 Q8GGU1 PRELIMINARY; PRT; 277 AA.
 ID AC Q8GGU1
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DE 28kDa outer membrane protein gene 17.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wakulla;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479838; AAO12957.1; -;
 SQ SEQUENCE 277 AA; 30294 MW; OCE7EDB51F2D854E CRC64;

Query Match 58.9%; Score 63; DB 2; Length 277;
 Best Local Similarity 52.6%; Pred. No. 0.069;
 Matches 10; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVGFLKQKQNDGSAIS 19
 DB 60 NATVALYGLKQDWNNGASAS 78

RESULT 31
 Q8G8W7 PRELIMINARY; PRT; 277 AA.
 ID AC Q8G8W7
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DE 28kDa outer membrane protein gene 17.
 OS Ehrlichia chaffeensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Anaplasmataceae; Ehrlichia.
 OX NCBI_TaxID=945;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=West Paces, Heartland, and St. Vincent;
 RX PubMed=12496165;
 RA Cheng C., Paddock C.D., Ganta R.R.;
 RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
 by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
 and Other Regions of the Genome.";
 RL Infect. Immun. 71:187-195(2003).
 DR EMBL; AF479835; AAO12942.1; -;
 DR EMBL; AF479836; AAO12947.1; -;
 DR EMBL; AF479837; AAO12952.1; -;
 SQ SEQUENCE 277 AA; 30276 MW; 4CF746150FF63FF4 CRC64;

Query Match 58.9%; Score 63; DB 2; Length 277;

```
Best Local Similarity 52.6%; Pred. No. 0.069;
Matches 10; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTVGVFLKQKQNDGSAIS 19
   | | | | | | | | | | | | | | |
Db 60 NATVALYGLKQDNGASAS 78

RESULT 32
ID O85360 PRELIMINARY; PRT; 133 AA.
AC O85360;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 28 kDa major surface antigen-2 (Fragment).
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,
  Allenan A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
  the tribe Ehrlichiae."; Commun. 247:636-643(1998).
RL Blochem. Biophys. Res. Commun. 247:636-643(1998).
DR EMBL; AF062762; AAC26722.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
FT NON_TER 133
SQ SEQUENCE 133 AA; 14903 MW; 9E422CAAF3FB91AC CRC64;

Query Match 57.5%; Score 61.5; DB 2; Length 133;
Best Local Similarity 58.4%; Pred. No. 0.053;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 TTVGVFLKQKQNDGSAISN 20
   ||| | | | | | | | | | |
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 33
ID Q9F474 PRELIMINARY; PRT; 283 AA.
AC Q9F474;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE P28-5 (Major outer membrane protein P30-3).
GN P28-5 OR P30-3.
OS Ehrlichia canis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=944;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=99242757; PubMed=10225842;
RA McBride J.W., Yu, Xj, Walker D.H.;
RT "Molecular cloning of the gene for a conserved major immunoreactive
  28-kilodalton protein of Ehrlichia canis: a potential serodiagnostic
  antigen.";
RL Clin. Diagn. Lab. Immunol. 6:392-399(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Jake;
RX MEDLINE=20432107; PubMed=10974556;
RA McBride J.W., Yu X.J., Walker D.H.;
RT "A conserved, transcriptionally active p28 multigene locus of
  Ehrlichia canis.";
```

```
Gene 254:245-252(2000).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=98371112; PubMed=9705412;
RA Ohashi N., Unver A., Zhi N., Rikihisa Y.;
RT "Cloning and characterization of multigenes encoding the
  immunodominant 30-kilodalton major outer membrane proteins of
  Ehrlichia canis and application of the recombinant protein for
  serodiagnosis.";
RL J. Clin. Microbiol. 36:2671-2680(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Oklahoma;
RX MEDLINE=21153566; PubMed=11254561;
RA Ohashi N., Rikihisa Y., Unver A.;
RT "Analysis of transcriptionally Active Gene Clusters of Major Outer
  Membrane Protein Multigene Family in Ehrlichia canis and E.
  chaffeensis.";
RL Infect. Immun. 69:2083-2091(2001).
DR EMBL; AF082744; AAG14360.1; -
DR EMBL; AF078553; AAK28698.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 283 AA; 31626 MW; 24A4B237C9C10715 CRC64;

Query Match 57.5%; Score 61.5; DB 2; Length 283;
Best Local Similarity 68.4%; Pred. No. 0.12;
Matches 13; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 2 TTVGVFLKQKQNDGSAISN 20
   ||| | | | | | | | | | |
Db 62 TTV-VYGLKENWAGDAISS 79

RESULT 34
ID O52106 PRELIMINARY; PRT; 278 AA.
AC O52106;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE OMP-1E (28kDa outer membrane protein gene 17).
GN OMP-1E.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Ehrlichia.
OX NCBI_TaxID=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98084465; PubMed=9423849;
RA Ohashi N., Zhi N., Zhang Y., Rikihisa Y.;
RT "Immunodominant major outer membrane proteins of Ehrlichia chaffeensis
  are encoded by a polymorphic multigene family.";
RL Infect. Immun. 66:132-139(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARKANSAS;
RX MEDLINE=98321180; PubMed=9647746;
RA Reddy G.R., Sulsona C.R., Barbet A.F., Mahan S.M., BurrIDGE M.J.,
  Allenan A.R.;
RT "Molecular characterization of a 28 kDa surface antigen gene family of
  the tribe Ehrlichiae.";
RL Biochem. Biophys. Res. Commun. 247:636-643(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Osciola;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
  by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
  and Other Regions of the Genome.";
```

RL Infect. Immun. 71:187-195(2003).
DR EMBL; U72291; AAC02939.1; -
DR EMBL; AF479833; AAC26719.1; -
DR EMBL; AF479834; AAC12936.1; -
DR InterPro; IPR002566; Surface_Ag_msp4.
DR Pfam; PF01617; Surface_Ag_2; 1.
SQ SEQUENCE 278 AA; 30543 MW; E321E3CA259B87FD CRC64;

Query Match 56.1%; Score 60; DB 2; Length 278;
Best Local Similarity 50.0%; Pred. No. 0.21; 4; Indels 0; Gaps 0;
Matches 10; Conservative

QY 1 NTTGVFGLKQNWGSAISN 20
| | | : | | | : | | : | |
DB 60 NPTVALYGLKQDWEGISSSS 79

RESULT 35

Q8G8Q5 PRELIMINARY; PRT; 278 AA.

AC Q8G8Q5; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 28kDa outer membrane protein gene 17.
OS Ehrlichia chaffeensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Ehrlichia.
OX NCBI_taxid=945;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Liberty, and Jax;
RX PubMed=12496165;
RA Cheng C., Paddock C.D., Ganta R.R.;
RT "Molecular Heterogeneity of Ehrlichia chaffeensis Isolates Determined
RT by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes
RT and Other Regions of the Genome."
RL Infect. Immun. 71:187-195(2003).
DR EMBL; AF479833; AAC12963.1; -
DR EMBL; AF479840; AAC12969.1; -
SQ SEQUENCE 278 AA; 30381 MW; 7626778E5798D6B7 CRC64;

Query Match

Best Local Similarity 56.1%; Score 60; DB 2; Length 278;
Matches 10; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQNWGSAISN 20
| | | : | | | : | | : | |
DB 60 NPTVALYGLKQDWEGISSSS 79

Search completed: October 6, 2003, 07:52:34
Job time : 30.1642 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 : Search time 10.7463 Seconds
(without alignments)
178.981 Million cell updates/sec

Title: US-09-765-739a-2

Perfect score: 110

Sequence: 1 NPTTGVFLKQWDGATIKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	90	81.8	280	2	JE0217	28k surface antige
2	82	74.5	276	2	JE0218	28k surface antige
3	71	64.5	286	2	JE0219	28k surface antige
4	59	53.6	278	2	JE0216	major antigenic pr
5	59	53.6	284	2	I40882	probable RTX famil
6	52	47.3	5188	2	B85547	hypothetical prote
7	52	47.3	5291	2	F90696	hypothetical prote
8	51	46.4	1191	2	A53491	bumetanide-sensiti
9	49	44.5	133	2	JE0221	28k surface antige
10	49	44.5	634	2	F97172	flagellar hook-ass
11	47	42.7	540	2	S54586	probable membrane
12	47	42.7	584	2	C48658	flagellin - Escher
13	46.5	42.3	267	2	C83242	conserved hypotet
14	46	41.8	261	2	G84057	hypothetical prote
15	45	40.9	160	2	A75466	2-demethylmenaquin
16	45	40.9	756	2	T20109	hypothetical prote
17	44	40.0	281	2	AH3012	outer surface prot
18	44	40.0	284	2	B69945	phage-related prot
19	44	40.0	284	2	G98271	hypothetical prote
20	44	40.0	465	1	S47738	cytochrome-c perox
21	44	40.0	465	2	F91178	cytochrome-c perox
22	44	40.0	465	2	G86024	cytochrome-c perox
23	44	40.0	534	2	C82096	aminoacyl-histidin
24	44	40.0	648	1	P3BPF6	p3 protein - phage
25	44	40.0	1004	2	JH0470	Na+/K+-exchanging
26	43.5	39.5	290	2	S78787	hypothetical prote
27	43	39.1	280	2	D70976	hypothetical prote
28	43	39.1	427	2	I40167	dihydroorotase (EC
29	43	39.1	470	2	JC4098	tetracycline 6-hyd

30	43	39.1	482	2	G75483	probable leucyl am
31	43	39.1	681	2	G87276	hypothetical prote
32	43	39.1	1037	2	A56594	Na+/K+-exchanging
33	43	39.1	1849	2	C88822	hypothetical prote
34	43	39.1	1944	2	AH3098	rhizoblocin/RTX to
35	43	39.1	1990	2	A96188	probable phosphos
36	42.5	38.6	123	2	AE0293	conserved hypotet
37	42	38.2	160	2	AI2476	hypothetical prote
38	42	38.2	224	2	C72390	hypothetical prote
39	42	38.2	287	1	S56603	probable formate a
40	42	38.2	287	2	A91296	probable activatin
41	42	38.2	307	2	D86137	probable activatin
42	42	38.2	328	2	T44893	hypothetical prote
43	42	38.2	362	2	F75379	S-adenosylmethioni
44	42	38.2	413	2	T08297	conserved hypotet
45	42	38.2	556	2	A41870	dnas protein - Str
46	42	38.2	725	2	A34474	hypothetical prote
47	42	38.2	1020	2	A04492	Na+/K+-exchanging
48	41.5	37.7	176	2	B64549	conserved hypotet
49	41.5	37.7	293	2	A84110	sugar ABC transpor
50	41.5	37.7	366	2	A81183	glycine cleavage s
51	41.5	37.7	368	2	G81919	probable aminometh
52	41.5	37.7	468	2	T23091	hypothetical prote
53	41.5	37.7	622	2	AB1090	bacteriophage mino
54	41	37.3	95	2	E90774	hypothetical prote
55	41	37.3	95	2	B85637	hypothetical prote
56	41	37.3	130	2	C88102	protein W09G10.6 {
57	41	37.3	244	2	E84885	hypothetical prote
58	41	37.3	269	2	A69997	hypothetical prote
59	41	37.3	328	2	AB1580	N-acetylmuramoyl-L
60	41	37.3	376	2	F53381	peptide ABC transp
61	41	37.3	377	2	F69008	acetyltransferase
62	41	37.3	389	2	T43979	hypothetical prote
63	41	37.3	395	1	A35629	mevalonate kinase
64	41	37.3	412	2	T09313	immediate-early pr
65	41	37.3	466	2	AB0960	probable cytochrom
66	41	37.3	478	2	AB0779	probable lipoprote
67	41	37.3	489	2	AF3594	sensory transducti
68	41	37.3	653	2	F85620	partial fibrial u
69	41	37.3	653	2	H90756	partial fibrial u
70	41	37.3	749	2	A45687	outer capsid prote
71	41	37.3	781	2	G96991	secreted protease
72	41	37.3	850	2	H82886	topoisomerase IV s
73	41	37.3	866	1	C64834	probable outer mem
74	41	37.3	1069	2	A81050	exodeoxyribonuclea
75	40.5	36.8	1025	2	A54718	dihydropyrimidine
76	40	36.4	84	2	T08232	probable gas-vesic
77	40	36.4	84	2	JQ1130	gas-vesicle protei
78	40	36.4	85	2	T27307	hypothetical prote
79	40	36.4	120	2	B69971	conserved hypotet
80	40	36.4	282	2	F64695	biotin synthetase
81	40	36.4	282	2	H71823	biotin synthetase
82	40	36.4	292	2	G84092	phage-related prot
83	40	36.4	299	2	T23932	hypothetical prote
84	40	36.4	328	2	AH1226	N-acetylmuramoyl-L
85	40	36.4	341	2	S73685	hypothetical prote
86	40	36.4	361	2	T39723	GTPase activating
87	40	36.4	408	2	S78830	hypothetical prote
88	40	36.4	489	2	T19227	queuine tRNA-ribos
89	40	36.4	490	2	G85354	hypothetical prote
90	40	36.4	500	2	F70012	leucyl aminopeptid
91	40	36.4	585	2	F90961	flagellin [importe
92	40	36.4	585	2	F85809	hypothetical prote
93	40	36.4	710	2	C98235	probable hydroxama
94	40	36.4	716	2	A13050	hydroxamate-type f
95	40	36.4	820	2	D83337	probable TonB-depe
96	40	36.4	964	1	VCLJC6	env polyprotein pr
97	40	36.4	1020	2	B24639	Na+/K+-exchanging
98	40	36.4	1023	1	A24414	Na+/K+-exchanging
99	40	36.4	1038	1	S03632	Na+/K+-exchanging
100	40	36.4	1205	2	A55015	bumetanide-sensiti

ALIGNMENTS

```
RESULT 1
JE0217
28k surface antigen 4 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0217
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burr ridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0217
A:Molecule type: DNA
A:Residues: 1-280 <RED>
A:Cross-references: GB:AF062761

Query Match      81.8%; Score 90; DB 2; Length 280;
Best Local Similarity 88.9%; Pred. NO. 1.8e-06;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDGATI 18
   ||| |||||:||||:|
Db 60 NTTGVFGLKQDWDGSTI 77

RESULT 2
JE0218
28k surface antigen 5 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0218
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burr ridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0218
A:Molecule type: DNA
A:Residues: 1-276 <RED>
A:Cross-references: GB:AF062761

Query Match      74.5%; Score 82; DB 2; Length 276;
Best Local Similarity 70.0%; Pred. NO. 3.1e-05;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDGATIKD 20
   ||| |||||:||||:|
Db 59 NTTGVFGLKQNWGSAISN 78

RESULT 3
JE0219
28k surface antigen 2 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0219
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burr ridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0219
A:Molecule type: DNA
A:Residues: 1-286 <RED>
A:Cross-references: GB:AF062761

Query Match      64.5%; Score 71; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. NO. 0.0017;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 1 NTTGVFGLKQDWDGATI 18
   ||| |||||:||||:|
Db 60 NTTGVFGLKQDWDRCVI 77

RESULT 4
JE0216
28k surface antigen 3 - Ehrlichia chaffensis
N:Alternate names: MAP1
C:Species: Ehrlichia chaffensis
C>Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 17-Mar-1999
C:Accession: JE0216
R:Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burr ridge, M.J.; Alleman, A.R.
Biochem. Biophys. Res. Commun. 247, 636-643, 1998
A:Title: Molecular characterization of a 28kDa surface antigen gene family of the tribe
A:Reference number: JE0216; MUID:98321180; PMID:9647746
A:Accession: JE0216
A:Molecule type: DNA
A:Residues: 1-278 <RED>
A:Cross-references: GB:AF062761

Query Match      53.6%; Score 59; DB 2; Length 278;
Best Local Similarity 60.0%; Pred. NO. 0.12;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 NTTGVFGLKQDWDG 15
   ||| |||||:|
Db 60 NPTVALYGLKQDWDG 74

RESULT 5
I40882
major antigenic protein - heartwater rickettsia
C:Species: Cowdria ruminantium (heartwater rickettsia)
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 08-Oct-1999
C:Accession: I40882; S42827
R:van Vliet, A.H.; Jongejans, F.; van Kleef, M.; van der Zeijst, B.A.
Infect. Immun. 62, 1451-1456, 1994
A:Title: Molecular cloning, sequence analysis, and expression of the gene encoding the
A:Reference number: I40882; MUID:94178956; PMID:8132352
A:Accession: I40882
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-284 <RES>
A:Cross-references: EMBL:X74250; NID:g454266; PIDN:CAA52309.1; PID:g454267
C:Genetics:
A:Gene: map1

Query Match      53.6%; Score 59; DB 2; Length 284;
Best Local Similarity 76.9%; Pred. NO. 0.13;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTGVFGLKQDWDG 15
   ||| |||||:|
Db 63 TKAVFGLKKDWDG 75

RESULT 6
B85547
probable RTX family exoprotein [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85547
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85547
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5188 <STO>
A:Cross-references: GB:AE005174; NID:g12513368; PIDN:AAG54838.1; GSPDB:GN00145; UWGP:
```

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: Z0615

Query Match 47.3%; Score 52; DB 2; Length 5188;

Best Local Similarity 60.0%; Pred. No. 37;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTGVFGLKQDWGCA 16

||||| : |||||

Db 4803 TTSGVAAMDYDWDGA.4817

RESULT 7

F90696

hypothetical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain R)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C;Accession: F90696

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: F90696

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-5291 <HAY>

A;Cross-references: GB:BA000007; PIDN:BA33965.1; PID:gl13360000; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RMD 0509952

C;Genetics:

A;Gene: ECs0542

Query Match 47.3%; Score 52; DB 2; Length 5291;

Best Local Similarity 60.0%; Pred. No. 38;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTGVFGLKQDWGCA 16

||||| : |||||

Db 4906 TTSGVAAMDYDWDGA 4920

RESULT 8

A53491

bumetanide-sensitive Na-K-Cl cotransporter - spiny dogfish

C;Species: Squalus acanthias (spiny dogfish)

C;Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 02-Mar-2001

C;Accession: A53491

R;Xu, J.C.; Lytle, C.; Zhu, T.T.; Payne, J.A.; Benz Jr., E.; Forbush III, B.

Proc. Natl. Acad. Sci. U.S.A. 91, 2201-2205, 1994

A;Title: Molecular cloning and functional expression of the bumetanide-sensitive Na-K-Cl

A;Reference number: A53491; MUID:94181560; PMID:8134373

A;Accession: A53491

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1191 <XUA>

A;Cross-references: GB:U05958; NID:g454096; PIDN:AAB60617.1; PID:g454097

C;Superfamily: rat bumetanide-sensitive Na+/K+/Cl-cotransport protein

Query Match 46.4%; Score 51; DB 2; Length 1191;

Best Local Similarity 58.8%; Pred. No. 11;

Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGVFGLKQDWGATIKD 20

||||| : |||||

Db 856 TLVFGPKDWRQALMKD 872

RESULT 9

JE0221

28k surface antigen 2 - Ehrlichia canis

C;Species: Ehrlichia canis

C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 21-Jul-2000

C;Accession: JE0221

R;Reddy, G.R.; Sulsona, C.R.; Barbet, A.F.; Mahan, S.M.; Burrige, M.J.; Alleman, A.

Biochem. Biophys. Res. Commun. 247, 636-643, 1998

A;Title: Molecular characterization of a 28kDa surface antigen gene family of the tri

A;Reference number: JE0216; MUID:98321180; PMID:9647746

A;Accession: JE0221

A;Molecule type: DNA

A;Residues: 1-133 <RED>

A;Cross-references: GB:AF062762; NID:g3327964; PIDN:AAC26722.1; PID:g3327966

Query Match 44.5%; Score 49; DB 2; Length 133;

Best Local Similarity 56.2%; Pred. No. 2;

Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 TTGVFGLKQDWGATYI 18

||||| : |||||

Db 62 TTVVYGLKENWAGDAI 77

RESULT 10

F97172

flagellar hook-associated protein FlgK [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C;Accession: F97172

R;Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; I

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: F97172

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-634 <KUR>

A;Cross-references: GB:AE001437; PIDN:AAK80169.1; PID:gl5025209; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC2212

Query Match 44.5%; Score 49; DB 2; Length 634;

Best Local Similarity 50.0%; Pred. No. 11;

Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 NTTTGVFGLKQDWGATIKD 20

:| ||| :| ||| :|

Db 536 DTDGTGNTIKSDPDGVTVD 555

RESULT 11

S54586

probable membrane protein YMR279c - yeast (Saccharomyces cerevisiae)

N;Alternate names: hypothetical protein YMR021.05c

C;Species: Saccharomyces cerevisiae

C;Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002

C;Accession: S54586

R;Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, May 1995

A;Reference number: S54582

A;Accession: S54586

A;Molecule type: DNA

A;Residues: 1-540 <PEA>

A;Cross-references: EMBL:Z49704; NID:g825540; PIDN:CAA89777.1; PID:g825545; GSPDB:GN

A;Experimental source: strain AB972

C;Genetics:

A;Gene: MIPS:YMR279c

A;Cross-references: SGD:S0004892

A;Map position: 13R

C;Superfamily: aminotriazole resistance protein YML116w

C;Keywords: transmembrane protein

F;109-125/Domain: transmembrane #status predicted <TM1>

F;139-155/Domain: transmembrane #status predicted <TM2>

F;174-190/Domain: transmembrane #status predicted <TM3>

F;233-249/Domain: transmembrane #status predicted <TM4>

```

RESULT 14
G84057
hypothetical protein BH3263 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G84057
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: G84057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-261 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:gi10175792; PIDN:BAE06982.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3263
C:Superfamily: Bacillus subtilis hypothetical protein ytmp

Query Match 41.8%; Score 46; DB 2; Length 261;
Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 12 DWDGATIKD 20
      |||||:|
Db 177 DWDGATVAD 185

RESULT 15
A75466
2-demethylmenaquinone 2-C-methyltransferase (EC 2.1.1.-) DR0859 [similarity] - Deinoc
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Sep-2000
C:Accession: A75466
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: A75466
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <WHI>
A:Cross-references: GB:AE001940; GB:AE000513; NID:g6458577; PIDN:AAF10437.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0859
A:Map position: 1
C:Keywords: methyltransferase

Query Match 40.9%; Score 45; DB 2; Length 160;
Best Local Similarity 50.0%; Pred. No. 10;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GVFGKQDWGATI 18
      |||||:|:|
Db 80 GVFGVNGWEGVII 93

RESULT 16
T20109
hypothetical protein C50F4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 12-May-2003
C:Accession: T20109
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z19225
A:Accession: T20109
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

```

A:Residues: 1-756 <WIL>
A:Cross-references: EMBL:270750; PIDN:CAA94737.1; GSPDB:GN00023; CESP:C50F4.2
A:Experimental source: clone C50F4
C:Genetics:
A:Gene: CESP:C50F4.2
A:Map position: 5
A:Introns: 24/1; 245/3; 315/2; 389/3; 537/3; 568/2; 747/1
C:Superfamily: 6-phosphofructokinase, eukaryotic type; 6-phosphofructokinase 1 homology

Query Match 40.9%; Score 45; DB 2; Length 756;
Best Local Similarity 53.3%; Pred. No. 57;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 6 VFGLKQDWGATIKD 20
| | | | | | | |
Db 436 VIGIKHWGDLKND 450
| | | | | | | |

RESULT 17
AH3012
outer surface protein Atu3708 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH3012
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, S.; Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH3012
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-281 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL44518.1; PID:g1742129; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3708
A:Map position: linear chromosome

Query Match 40.0%; Score 44; DB 2; Length 281;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 GVFLKQDWGATIKD 20
| | | | | | | |
Db 68 GPGYLQDFDTATIKD 83
| | | | | | | |

RESULT 18
B69945
phage-related protein homolog yqak - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: B69945
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berterio, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chao, A.; Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, Y. M.; Ogawa, K.; Ogiwara, B.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron, akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: B69945
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-284 <KUN>
A:Cross-references: GB:Z99117; GB:AL009126; NID:g2634966; PIDN:CAB14569.1; PID:g26350
A:Experimental source: strain 168
C:Genetics:
A:Gene: yqak
C:Superfamily: Escherichia coli recT protein

Query Match 40.0%; Score 44; DB 2; Length 284;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 7 FGLKQDWGATIK 19
| | | | | | | |
Db 193 FGKNDWDAMALK 205
| | | | | | | |

RESULT 19
G98271
hypothetical protein AGR_L_2258 [imported] - Agrobacterium tumefaciens (strain C58, C)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: G98271
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldn, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G98271
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-284 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89697.1; PID:g15159605; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2258
A:Map position: linear chromosome

Query Match 40.0%; Score 44; DB 2; Length 284;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 GVFLKQDWGATIKD 20
| | | | | | | |
Db 71 GPGYLQDFDTATIKD 86
| | | | | | | |

RESULT 20
S47738
cytochrome-c peroxidase (EC 1.11.1.5) - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 27-Jan-1995 #sequence_revision 02-Aug-1996 #text_change 01-Mar-2002
C:Accession: S47738; A65150
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47738
A:Molecule type: DNA
A:Residues: 1-465 <PLU>
A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AAB18494.1; PID:g466655
A:Note: this sequence contains three cytochrome c-type heme-binding motifs
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A65150
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-465 <BLAT>
A:Cross-references: GB:AE000428; GB:U00096; NID:g1789931; PIDN:AAC76543.1; PID:g17899
A:Experimental source: strain K-12, substrain MG1655

A:Gene: yhjA
C:Superfamily: Escherichia coli cytochrome-c peroxidase; Pseudomonas cytochrome-c peroxidase
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; oxidoreductase
F:181,426/Domain: Pseudomonas cytochrome-c peroxidase homology <PCC>
F:59,62/Binding site: heme (Cys) (covalent) #status predicted
F:63,195/Binding site: heme iron (His) (axial ligands) #status predicted
F:207,210/Binding site: heme iron (His) (covalent) (low potential) #status predicted
F:211,415/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
F:351,354/Binding site: heme (Cys) (covalent) (high potential) #status predicted
F:355,429/Binding site: heme iron (His, Met) (axial ligands) (high potential) #status predicted

Query Match 40.0%; Score 44; DB 1; Length 465;
Best Local Similarity 52.9%; Pred. No. 48;
Matches 9; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 6 VFGLKQDWDG--ATIKD 20
|| :||| ||| :|||
Db 243 VFNVEQFWDGRAATLQD 259

RESULT 21
F91178
cytochrome-c peroxidase (EC 1.11.1.5) [similarity] - Escherichia coli (strain O157:H7, S
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Sep-2001
C:Accession: F91178
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A95629; MUID:21156231; PMID:11258796
A:Accession: F91178
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <HAY>
A:Cross-references: GB:BA00007; PIDN:BA37821.1; PID:g13363872; GSPDB:GN00154
A:Experimental source: strain O157:H7, Substrain RMD 050952
C:Genetics:
A:Gene: ECs4398
C:Superfamily: Escherichia coli cytochrome-c peroxidase; Pseudomonas cytochrome-c peroxi
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:59,62/Binding site: heme (Cys) (covalent) (high potential) #status predicted
F:207,210/Binding site: heme (Cys) (covalent) (low potential) #status predicted
F:211,415/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
F:351,354/Binding site: heme (Cys) (covalent) (high potential) #status predicted
F:355,429/Binding site: heme iron (His, Met) (axial ligands) (high potential) #status pr

Query Match 40.0%; Score 44; DB 2; Length 465;
Best Local Similarity 52.9%; Pred. No. 48;
Matches 9; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 6 VFGLKQDWDG--ATIKD 20
|| :||| ||| :|||
Db 243 VFNVEQFWDGRAATLQD 259

RESULT 22
G86024
cytochrome-c peroxidase (EC 1.11.1.5) [similarity] - Escherichia coli (strain O157:H7, S
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G86024
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G86024
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-465 <STO>
A:Cross-references: GB:AE005174; NID:g12518230; PIDN:AAG58659.1; GSPDB:GN00145; UWGP:249

A:Experimental source: strain O157:H7, Substrain EDL933
C:Genetics:
A:Gene: yhjA
C:Superfamily: Escherichia coli cytochrome-c peroxidase; Pseudomonas cytochrome-c per
C:Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase
F:59,62/Binding site: heme (Cys) (covalent) (high potential) #status predicted
F:207,210/Binding site: heme (Cys) (covalent) (low potential) #status predicted
F:211,415/Binding site: heme iron (His) (axial ligands) (low potential) #status predicted
F:351,354/Binding site: heme (Cys) (covalent) (high potential) #status predicted
F:355,429/Binding site: heme iron (His, Met) (axial ligands) (high potential) #status

Query Match 40.0%; Score 44; DB 2; Length 465;
Best Local Similarity 52.9%; Pred. No. 48;
Matches 9; Conservative 4; Mismatches 2; Indels 2; Gaps 1;

QY 6 VFGLKQDWDG--ATIKD 20
|| :||| ||| :|||
Db 243 VFNVEQFWDGRAATLQD 259

RESULT 23
C82096
aminocycl-histidine dipeptidase VC2279 [imported] - Vibrio cholerae (strain N16961 se
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82096
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: C82096
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <HE>
A:Cross-references: GB:AE004299; GB:AE003852; NID:g9656835; PIDN:AAF95423.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2279
A:Map position: 1

Query Match 40.0%; Score 44; DB 2; Length 534;
Best Local Similarity 70.0%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 TGVFGLKQDW 13
|| ||||| :
Db 198 TGAFLKKEGW 207

RESULT 24
P3BPF6
P3 protein - phage phi-6
C:Species: phage phi-6
A:Note: host Pseudomonas phaseolocola
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 23-Jul-1999
C:Accession: C28648
R:Gottlieb, P.; Metzger, S.; Romantschuk, M.; Carton, J.; Strassman, J.; Bamford, D.H
Virology 163, 183-190, 1988
A:Title: Nucleotide sequence of the middle dsRNA segment of bacteriophage phi-6: plac
A:Reference number: A94372; MUID:88160044; PMID:3347997
A:Accession: C28648
A:Molecule type: genomic RNA
A:Residues: 1-648 <GOT>
A:Cross-references: GB:M17462; NID:g862621; PIDN:AAA68485.1; PID:g215490
C:Comment: The genome of this phage consists of three segments of double-stranded RNA
C:Comment: This protein is required for adsorption onto host cells.
C:Genetics:
A:Gene: 3
A:Map position: segment M
A:Superfamily: phage phi-6 p3 protein
C:Keywords: late protein

с, субстратом. субстратом субстратом, субстратом субстратом

RESULT 29

JC4098
tetracycline 6-hydroxylase - Streptomyces aureofaciens
C:Species: Streptomyces aureofaciens
C>Date: 23-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 13-Sep-1998
C:Accession: JC4098
R:Dairi, T.; Nakano, T.; Aisaka, K.; Katsumata, R.; Hasegawa, M.
BioSci. Biotechnol. Biochem. 59, 1099-1106, 1995
A:Title: Cloning and nucleotide sequence of the gene responsible for chlorination of tetracycline
A:Reference number: JC4098; MUID:95337551; PMID:7612997
A:Accession: JC4098
A:Molecule type: DNA
A:Residues: 1-470 <DAI>
A:Cross-references: DDBJ:D38214
A:Experimental source: NRRL3203
C:Genetics:
A:Gene: chl-2
C:Superfamily: tetracycline 6-hydroxylase

Query Match 39.1%; Score 43; DB 2; Length 470;
Best Local Similarity 50.0%; Pred. No. 69;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 TGVEGLKQDWDGATIK 19
: || | || || :
Db 119 SGVTGFAQADGVTVE 134

RESULT 30

G75483
probable leucyl aminopeptidase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Mar-2003
C:Accession: G75483
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999.
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: G75483
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-482 <WHI>
A:Cross-references: GB:AE001928; GB:AE000513; NID:g6458421; PIDN:AAF10295.1; PID:g645842
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0717
A:Map position: 1
C:Superfamily: Cytosol aminopeptidase

Query Match 39.1%; Score 43; DB 2; Length 482;
Best Local Similarity 46.7%; Pred. No. 71;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTGVEGLKQDWDGA 16
: |:::| |::
Db 266 TSGMYGMKNDMGA 280

RESULT 31

G87276
hypothetical protein CC0224 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: G87276
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: G87276

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-681 <STO>
A:Cross-references: GB:AE005673; NID:gl3421351; PIDN:AAK22211.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0224

Query Match 39.1%; Score 43; DB 2; Length 681;
Best Local Similarity 44.4%; Pred. No. 1e+02;
Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 3 TTGVEGLKQDWDGATIK 20
: | | | | |::
Db 543 TGGRETFKNDWNETHVKD 560

RESULT 32

A56594
Na+/K+-exchanging ATPase (EC 3.6.3.9) alpha chain - cat flea
N:Alternate names: sodium pump alpha subunit
C:Species: Ctenocephalides felis (cat flea)
C>Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 19-Apr-2002
C:Accession: A56594
R:Reeves, S.A.; Yamanaka, M.K.
Insect Biochem. Mol. Biol. 23, 809-814, 1993
A:Title: Cloning and sequence analysis of the alpha subunit of the cat flea sodium pu
A:Reference number: A56594; MUID:93386186; PMID:8397036
A:Accession: A56594
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1037 <REE>
A:Cross-references: GB:S66043; NID:g432541; PIDN:AA828239.1; PID:g432542
A:Note: sequence extracted from NCBI backbone (NCBI:137976, NCBI:137977)
C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding doma
C:Keywords: ATP; glycoprotein; hydrolyase; phosphoprotein; potassium transport; sodium
F:601-797/Domain: ATPase nucleotide-binding domain; homology <ATM>
F:601-797/Active site: Asb (aspartylphosphate intermediate) #status predicted
F:197/Binding site: carbonylate (Asn) (covalent) #status predicted
F:522/Binding site: ATP (lys) #status predicted

Query Match 39.1%; Score 43; DB 2; Length 1037;
Best Local Similarity 33.3%; Pred. No. 1.6e+02;
Matches 5; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 6 VFGLKQDWDGATIK 20
: ||::: | | :
Db 897 LFGIRKQWDSKAVND 911

RESULT 33

C86822
hypothetical protein yqk [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86822
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86822
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1649 <STO>
A:Cross-references: GB:AE005176; PID:gl2724583; PIDN:AAK05677.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yqk

Query Match 39.1%; Score 43; DB 2; Length 1649;
Best Local Similarity 46.7%; Pred. No. 2.7e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 3 TTGVEGLKQDWDGAT 17

```
|||||:
Db      1603 TTGTFKLQNNWNSTS 1617

RESULT 34
AH3098
rhizobiocin/RTX toxin [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AH3098
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AH3098
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1944 <KUR>
A;Cross-references: GB:AE008689; PIDN:AAL45206.1; PID:g17742885; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: rzca
A;Map position: linear chromosome

Query Match      39.1%; Score 43; DB 2; Length 1944;
Best Local Similarity 44.4%; Pred. No. 3.2e+02;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      3 TTGVFGLKQDWGATIKD 20
      |||:  :||| |
Db      1077 TTGLMSRELTWNGATNSD 1094

RESULT 35
A96188
probable phosphoesterase (EC 3.1.-.-) yvNB [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: A96188
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: A96188
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1990 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89027.1; PID:g15158819; GSPDB:GN00170
C;Genetics:
A;Gene: AGK_L_909
A;Map position: linear chromosome

Query Match      39.1%; Score 43; DB 2; Length 1990;
Best Local Similarity 44.4%; Pred. No. 3.3e+02;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      3 TTGVFGLKQDWGATIKD 20
      |||:  :||| |
Db      1123 TTGLMSRELTWNGATNSD 1140
```

Search completed: October 6, 2003, 07:49:24
Job time : 16.7463 secs

3 PAGE BLANK (USP)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:02 ; Search time 5.67164 Seconds
(without alignments)
165.831 Million cell updates/sec

Title: US-09-765-739A-2

Perfect score: 110

Sequence: 1 NTTTGVFLKQDWDGATIKD 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	46.4	1191	S122_SQUAC	P55013 squalus aca
2	47	42.7	540	YMBM_YEAST	Q03263 saccharomyc
3	45	40.9	160	MENG_DEIRA	Q9RW10 deinococcus
4	45	40.9	756	K6PF_CAEEL	Q27483 caenorhabdi
5	44	40.0	284	YQAK_BACSU	P45908 bacillus su
6	44	40.0	465	YHJA_ECOLI	P37197 escherichia
7	44	40.0	648	VP3_BPPH6	P11129 bacterioph
8	44	40.0	1004	AT1B_ARTSP	P28774 artemia san
9	43.5	39.5	290	AROE_SYNY3	P28774 artemia san
10	43	39.1	427	PYRC_BACCL	P74591 synecocyst
11	43	39.1	1021	ALAL_CANFA	P46538 bacillus ca
12	42	38.2	287	VJWV_ECOLI	P50997 canis fami
13	42	38.2	307	YML6_MYCLE	P39409 escherichia
14	42	38.2	362	QUEA_DEIRA	O32960 mycobacteri
15	42	38.2	656	DNAA_STRCO	Q9RU19 deinococcus
16	42	38.2	1020	ALIA2_HUMAN	P27902 streptomyc
17	41.5	37.7	366	GCST_NEIMA	P50993 homo sapien
18	41.5	37.7	368	GCST_NEIMA	Q9K018 neisseria m
19	41	37.3	377	YAE7_METHH	O9JVP2 neisseria m
20	41	37.3	395	KINE_RAT	O27139 methanobact
21	41	37.3	684	HTPG_PORGI	P17256 rattus norv
22	41	37.3	749	VP4_ROTGA	Q9S3Q2 porphyromon
23	41	37.3	866	YCBS_ECOLI	Q04916 rotavirus (
24	40.5	36.8	196	ANFL_CHICK	P75857 escherichia
25	40.5	36.8	734	PURL_ZYMO	P79775 gallus gall
26	40.5	36.8	1025	DPID_HUMAN	Q9RCQ6 zymomonas m
27	40	36.4	84	GVMI_HALNI	Q12882 homo sapien
28	40	36.4	120	YRAJ_BACSU	P24377 halobacteri
29	40	36.4	191	PGHD_URSAR	Q07934 bacillus su
30	40	36.4	282	BIOB_HELPJ	Q29562 ursus arcto
31	40	36.4	282	BIOB_HELPJ	Q25556 helicobacte
32	40	36.4	341	Y33B_MYCPN	Q25556 helicobacte
33	40	36.4	489	UBPE_CAEEL	Q17361 caenorhabdi

ALIGNMENTS

RESULT 1

34	40	36.4	500	1	AMPA_BACSU	O32106 bacillus su
35	40	36.4	523	1	GUAA_CORGL	Q8NSR1 corynebacte
36	40	36.4	942	1	ENV_CAEVG	P31627 caprine art
37	40	36.4	1020	1	ALIA2_RAT	P06886 rattus norv
38	40	36.4	1023	1	ALIA1_HUMAN	P05023 homo sapien
39	40	36.4	1041	1	ATNA_DROME	P13607 drosophila
40	40	36.4	1205	1	S122_HUMAN	P55012 mus musculu
41	40	36.4	1212	1	S122_HUMAN	P55011 homo sapien
42	40	36.4	1295	1	GLPI_CAEEL	P13508 caenorhabdi
43	40	36.4	1429	1	LI12_CAEEL	P14585 caenorhabdi
44	39	35.5	105	1	YGRM_MICEC	P24620 micromonosp
45	39	35.5	118	1	ANFD_RANCA	P40756 rana catesb
46	39	35.5	149	1	DTD_CLOAB	Q97922 clostridium
47	39	35.5	188	1	EFPL_XYLEF	Q87C43 xylella fas
48	39	35.5	189	1	EFPL_XYLEF	Q9PBE1 xylella fas
49	39	35.5	267	1	NUSL_ASPOR	P24021 aspergillus
50	39	35.5	275	1	APAH_PASMU	P57922 pasteurella
51	39	35.5	282	1	PUNA_CELSP	P81989 cellulomona
52	39	35.5	298	1	YMA2_MYCBO	Q02278 mycobacteri
53	39	35.5	326	1	GBLP_NICPL	P93340 nicotiana p
54	39	35.5	326	1	GBLP_TOBAC	P49026 nicotiana t
55	39	35.5	339	1	F16Q_BRANA	P46267 brassica na
56	39	35.5	356	1	MURC_YERPE	Q8Z1A9 yersinia pe
57	39	35.5	422	1	YF02_MYCPN	P75285 mycoplasma
58	39	35.5	428	1	YB01_MYCPN	P75568 mycoplasma
59	39	35.5	438	1	YCO5_MYCPN	P75571 mycoplasma
60	39	35.5	741	1	PLO3_MOUSE	Q9R0E1 mus musculu
61	39	35.5	802	1	XYND_RUMFL	Q53317 ruminococcu
62	39	35.5	896	1	POL_HTLIC	P14078 human t-cel
63	39	35.5	1021	1	ALAI_HORSE	P18907 equus cabal
64	39	35.5	1025	1	DPYD_BOVIN	Q28007 bos taurus
65	39	35.5	1025	1	DPYD_PIG	Q28943 sus scrofa
66	38.5	35.0	319	1	AES_ECOLI	P23872 escherichia
67	38.5	35.0	536	1	FYN_XIPHE	P27446 xiphophorus
68	38.5	35.0	567	1	UREL_PROMI	P17086 proteus mir
69	38.5	35.0	1182	1	CGA2_HELPY	P55746 helicobacte
70	38	34.5	105	1	YGRM_MICRO	P24621 micromonosp
71	38	34.5	135	1	FABE_BOVIN	P55052 bos taurus
72	38	34.5	135	1	FABE_HUMAN	Q01469 homo sapien
73	38	34.5	162	1	MEGI_VIBCH	Q9KPK1 vibrio chol
74	38	34.5	180	1	RBS_MARPA	O64416 marchantia
75	38	34.5	197	1	VG23_BPMD2	Q38362 mycobacteri
76	38	34.5	338	1	IPNS_CBPAC	P05189 cephalospor
77	38	34.5	341	1	EPG_STRRA	P29541 streptomyc
78	38	34.5	360	1	MRAY_NEIMA	Q9J3Z3 neisseria m
79	38	34.5	360	1	MRAY_NEIMA	Q9K0Y6 neisseria m
80	38	34.5	363	1	QUEA_BRUME	Q8YHB3 brucella me
81	38	34.5	371	1	Y02B_ARCFU	O30207 archaeoglob
82	38	34.5	372	1	RFNG_CHICK	O12972 gallus gall
83	38	34.5	389	1	G22E_DROME	P58953 drosophila
84	38	34.5	430	1	Y588_MPTJA	Q58008 methanococc
85	38	34.5	469	1	KE4_HUMAN	Q92504 homo sapien
86	38	34.5	473	1	XYLA_CLOSR	P48790 clostridium
87	38	34.5	524	1	GUAA_CORAM	O52831 corynebacte
88	38	34.5	526	1	AOFA_RAT	P21396 rattus norv
89	38	34.5	547	1	IF37_MOUSE	O70194 mus musculu
90	38	34.5	548	1	IF37_HUMAN	O15371 homo sapien
91	38	34.5	676	1	SP1L_COTJA	P23499 coturnix co
92	38	34.5	680	1	YH2X_SCHPO	O74343 schizosacch
93	38	34.5	689	1	SYGE_PASMU	P57905 pasteurella
94	38	34.5	708	1	EPGL_STRCO	P40173 streptomyc
95	38	34.5	741	1	YL19_ARCFU	O28161 archaeoglob
96	38	34.5	748	1	GUNC_PSEFL	P27033 pseudomonas
97	38	34.5	792	1	OSTA_XYLEA	Q9PFA1 xylella fas
98	38	34.5	812	1	PAED_ECOLI	P06970 escherichia
99	38	34.5	959	1	N100_YEAST	Q02629 saccharomyc
100	38	34.5	1010	1	ALIA3_CHICK	P24798 gallus gall

S122_SQUAC STANDARD; PRT: 1191 AA.
ID S122_SQUAC
AC P5013;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Solute carrier family 12 member 2 (Bumetanide-sensitive sodium-
potassium)-chloride cotransporter 1 (NA-K-CL symporter) (NKCC).
GN SLC12A2 OR NKCC1
OS Squalus acanthias (Spiny dogfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalia; Squaloidei; Squalidae; Squalus.
OX NCBI_TaxID=7797;
RN [1]
RP SEQUENCE FROM N.A., AND PHOSPHORYLATION SITES.
RC TISSUE-Rectal gland;
RX MEDLINE=94181560; PubMed=8134373;
RA Xu J.-C., Lytle C., Zhu T.T., Payne J.A., Benz E. Jr., Forbush B. III;
RT "Molecular cloning and functional expression of the
bumetanide-sensitive Na-K-Cl cotransporter."
RL Proc. Natl. Acad. Sci. U.S.A. 91:2201-2205(1994).
CC -!- FUNCTION: Electrically silent transporter system. Mediates sodium
and chloride reabsorption. Plays a vital role in the regulation of
ionic balance and cell volume.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in many tissues.
CC -!- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U05958; AAB60617.1; -
CC PIR: A53491; A53491;
CC InterPro: IPR002293; AA/rel.permeasel.
CC InterPro: IPR004842; KCL_cotranspt.
CC InterPro: IPR002443; NAKCL_cotranspt.
CC PRINTS: PR01207; NAKCLTRNSPT.
CC TIGRFAMs: TIGR00930; 2a30; 1.
CC Potassium transport; Ion transport; Sodium transport; Symport;
KW Potassium transport; Potassium; Transmembrane; Phosphorylation.
FT DOMAIN 1 257 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 282 302 POTENTIAL.
FT DOMAIN 303 339 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 340 360 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT DOMAIN 404 407 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 408 428 POTENTIAL.
FT TRANSMEM 459 479 POTENTIAL.
FT DOMAIN 480 496 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 497 517 POTENTIAL.
FT TRANSMEM 555 585 POTENTIAL.
FT DOMAIN 586 630 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 631 651 POTENTIAL.
FT TRANSMEM 652 672 POTENTIAL.
FT DOMAIN 673 689 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 690 710 POTENTIAL.
FT TRANSMEM 776 796 POTENTIAL.
FT DOMAIN 797 1191 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1 27 30 POLY-PRO.
FT DOMAIN 51 57 POLY-GLY.
FT DOMAIN 67 70 POLY-ALA.
FT DOMAIN 89 92 POLY-ALA.
FT DOMAIN 96 99 POLY-ALA.
FT DOMAIN 126 129 POLY-PRO.
FT CARBOHYD 523 523 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 535 535 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 546 546 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT MOD_RES 189 189 PHOSPHORYLATION.
FT MOD_RES 1114 1114 PHOSPHORYLATION.
SQ SEQUENCE 1191 AA; 129774 MW; BA7BB9815431500C CRC64;
Query Match 46.4%; Score 51; DB 1; Length 1191;
Best Local Similarity 58.8%; Pred. No. 2,6; Indels 0; Caps 0;
Matches 10; Conservative 2; Mismatches 5;
QY 4 TGVFGLKQDWGATIKD 20
| | | | | : | | | | : | | |
Db 856 TLVFGFKDWROALMKD 872
RESULT 2
YMBM_YEAST STANDARD; PRT: 540 AA.
ID YMBM_YEAST
AC Q03263;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 59.6 kDa protein in DSK2-CAT8 intergenic region.
GN YMR279C OR YMR021.05C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAINS=S288c / AB972;
RX PubMed=9169872;
RA Bowman S., Chumacher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor K., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagers K., Lye G., Moutle S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT The nucleotide sequence of Saccharomyces cerevisiae chromosome
XIII.
RT Nature 387:90-93(1997).
RL -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
AS THE DRUG RESISTANCE TRANSLOCASE FAMILY). DHA12 SUBFAMILY.
CC STRONG, TO YEAST ATRI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 249704; CAA89777.1; -
CC PIR: S54586; S54586.
CC SGD: S0004892; YMR279C.
CC InterPro: IPR007114; MFS.
CC Hypothetical protein; Transport; Transmembrane.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 103 129 POTENTIAL.
FT TRANSMEM 132 152 POTENTIAL.
FT TRANSMEM 170 190 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 430 450 POTENTIAL.
FT TRANSMEM 462 482 POTENTIAL.
FT TRANSMEM 503 523 POTENTIAL.
SQ SEQUENCE 540 AA; 59561 MW; 687D06CB0D70AF91 CRC64;
Query Match 42.7%; Score 47; DB 1; Length 540;
Best Local Similarity 38.9%; Pred. No. 5;

Matches 7; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 NNTTGVGLKQDWDGATI 18
| : | | | | | :
Db 256 NVTNTHGLSMDWTGSAL 273

RESULT 3

MENG_DEIRA
ID MENG_DEIRA STANDARD; PRT; 160 AA.
AC Q9RW10;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable S-adenosylmethionine:2-demethylmenaquinone methyltransferase
DE (EC 2.1.1.1)
DE MENG OR DR0859.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RL / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL";
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: CONVERYS DIMETHYLMENAQUINONE (DMK) TO MENAQUINONE (MK)
CC (By similarity).
CC -!- PATHWAY: Menaquinone biosynthesis; last step.
CC -!- SIMILARITY: BELONGS TO THE MENG FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; AE001940; AAF10437.1; -
DR PIR; A75466; A75466.
DR TIGR; DR0859; -

DR HAMAP; MF_00471; -; 1.
DR InterPro; IPR005493; Methyltransf_6.
DR Pfam; PF03737; Methyltransf_6; 1.
KW Menaquinone biosynthesis; Transferase; Methyltransferase;
KW Complete proteome.
SQ SSEQUENCE 160 AA; 16860 MW; 54D6F226CA19EC0E CRC64;

Query Match 40.9%; Score 45; DB 1; Length 160;
Best Local Similarity 50.0%; Pred. No. 2.9;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GVFGKQDWDGATI 18
| | | | | : | | |
Db 80 GVFGVNGWGVII 93

RESULT 4

K6PF CAEEL
ID K6PF CAEEL STANDARD; PRT; 756 AA.
AC Q27483;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable 6-phosphofructokinase (EC 2.7.1.11) (Phosphofructokinase)
DE (Phosphohexokinase).
GN C50F4.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA McMurray A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -!- PATHWAY: Key control step of glycolysis.
CC -!- SIMILARITY: Belongs to the phosphofructokinase family. Two domains
CC subfamily.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; Z70750; CAA94737.1; -
DR PIR; T20109; T20109.
DR HSSP; P00512; 3PFK.
DR WormPep; C50F4.2; CE05467.
DR InterPro; IPR000023; Pfrfructkinase.
DR Pfam; PF00365; PFK; 2.
DR PRINTS; PR00476; PFRCTKINASE.
DR ProDom; PD000707; Pfrfructkinase; 2.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
KW Kinase; Transferase; Glycolysis; Repeat.
SQ SEQUENCE 756 AA; 83301 MW; 26A89B801D286534 CRC64;

Query Match 40.9%; Score 45; DB 1; Length 756;

Best Local Similarity 53.3%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 6 VFGLKQDWDGATIKD 20
| | | | | | | | | |
Db 436 VIGIKHGWDLKKNKD 450

RESULT 5

YOAK_BACSU
ID YOAK_BACSU STANDARD; PRT; 284 AA.
AC P45908;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yqak.
GN YOAK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=95219086; PubMed=7704261;
RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;
RT "Complete nucleotide sequence of a skin element excised by DNA
RT rearrangement during sporulation in Bacillus subtilis.";
RL Microbiology 141:323-327(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,

RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes";
RL Microbiology 142:3103-3111(1996).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN:168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschini C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerthoff A., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Etian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarovic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rochelle B., Rapoport G., Rey M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sato T., Scanlan E., Takagi T., Tanaka T., Terpstra P., Tognoni A.,
RA Sorokini A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Vandenbol M., Vannier F., Vassarotti A.,
RA Tosato V., Uchiyama S., Wedler H., Wedler H., Weitzenecker T.,
RA Viari A., Wambuit R., Wedler H., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
RN [4]
RN IDENTIFICATION:
RP MEDLINE=96084975; PubMed=7489895;
RA Medigue C., Moszer I., Viari A., Danchin A.,
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
RT computer system prototype";
RL Gene 165:GC37-GC51(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D32216; BAA06925.1; -
DR EMBL: D84432; BAA12386.1; -
DR EMBL: 299117; CAB14569.1; -
DR PIR: B69945; B69945.
DR Subtilist: BG11262; yqak.
DR InterPro: IPR004590; Rect.
DR Pfam: PF03837; Rect; 1
DR TIGRfam: TIGR00616; rect; 1.
KW Hypothetical protein. Complete proteome.
SQ SEQUENCE 284 AA; 32170 MW; F255261D4692ADB7 CRC64;
Query Match 40.0%; Score 44; DB 1; Length 284;
Best Local Similarity 53.8%; Pred. No. 7.8;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 7 FGLKQDWGATIK 19
||| :||| :|

Db 193 FGKNDWDAMALX 205
RESULT 6
YHJA_ECOLI STANDARD; PRT; 465 AA.
ID YHJA_ECOLI
AC P37197; 1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PROBABLE CYTOCHROME C PEROXIDASE (EC 1.1.1.5).
GN YHJA OR B3518.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes";
RL Nucleic Acids Res. 22:2576-2585(1994).
CC -1- CATALYTIC ACTIVITY: 2 ferrocyclochrome c + H(2)O(2) - 2
CC ferrocyclochrome c + 2 H(2)O.
CC -1- PTM: BINDS 3 HEMES (POTENTIAL).
CC -1- SIMILARITY: HIGH TO P.AERUGINOSA CYTOCHROME C551 PEROXIDASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00039; AAB18494.1; -
DR EMBL: AE000428; AAC76543.1; -
DR PIR: S47738; S47738.
DR HSSP: P14532; IEB7.
DR EcoGene: EGI2244; yHJA.
DR InterPro: IPR003088; Cyt_C1.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR004852; CytCP_Maug.
DR Pfam: PF03150; CCP_Maug; 1.
DR PROSITE: PS00190; CYTOCHROME_C; 3.
KW Hypothetical protein; Oxidoreductase; Peroxidase; Heme;
KW Electron transport; Complete proteome.
FT BINDING 59 59 HEME 1 (COVALENT) (BY SIMILARITY).
FT BINDING 62 62 HEME 1 (COVALENT) (BY SIMILARITY).
FT METAL 63 63 IRON (HEME 1 PROXIMAL LIGAND)
FT METAL 63 63 (BY SIMILARITY).
FT BINDING 207 207 HEME 2 (COVALENT) (BY SIMILARITY).
FT BINDING 210 210 HEME 2 (COVALENT) (BY SIMILARITY).
FT METAL 211 211 IRON (HEME 2 PROXIMAL LIGAND)
FT METAL 211 211 (BY SIMILARITY).
FT BINDING 351 351 HEME 3 (COVALENT) (BY SIMILARITY).
FT BINDING 354 354 HEME 3 (COVALENT) (BY SIMILARITY).
FT METAL 355 355 IRON (HEME 3 PROXIMAL LIGAND) (BY
FT METAL 415 415 SIMILARITY).
FT METAL 415 415 IRON (HEME 2 DISTAL LIGAND) (BY
FT METAL 429 429 SIMILARITY).
FT METAL 429 429 IRON (HEME 3 DISTAL LIGAND) (BY
FT METAL 429 429 SIMILARITY).
SQ SEQUENCE 465 AA; 51570 MW; 9F494A698949E6DA CRC64;
Query Match 40.0%; Score 44; DB 1; Length 465;
Best Local Similarity 52.9%; Pred. No. 13;
Matches 9; Conservative 4; Mismatches 2; Indels 2; Gaps 1;
QY 6 VFGLKQDWG--ATIKD 20
||| :||| :||| :|

```
Db      243 VFVNEQFWDGRAATLQD 259
VP3_BPPH6
ID VP3_BPPH6          STANDARD;          PRT;          548 AA.
AC P11129;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DE P3 protein.
GN P3.
OS Bacteriophage phi-6.
OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
OX NCBI_TaxID=10879;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.
RX MEDLINE=88160044; PubMed=3347997;
RA Gottlieb P., Metzger S., Romantschuk M., Carton J., Strassman J.,
BA Bamford D.H., Kalkinen N., Mindich L.;
RT "Nucleotide sequence of the middle dsRNA segment of bacteriophage phi
RL Virology 163:183-190(1988).
CC -!- FUNCTION: P3 PROTEIN IS NECESSARY FOR ADSORPTION ONTO HOST
CELLS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M17462; AAA68485.1; -.
CC PIR: C28648; P3BPP6.
CC KW Envelope protein.
CC SQ SEQUENCE 648 AA; 69178 MW; B188DFE02ACC54E3 CRC64;

Query Match 40.0%; Score 44; DB 1; Length 648;
Best Local Similarity 41.2%; Pred. No. 19;
Matches 7; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 6 VFG--LKQDWGATIKD 20
   |||: |||: |
Db 181 IFGWYKMDWEGSAVD 197

RESULT 8
ATIB_ARTSF
ID ATIB_ARTSF          STANDARD;          PRT;          1004 AA.
AC P28774;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sodium/potassium-transporting ATPase alpha chain (EC 3.6.3.9) (Sodium
DE pump) (Na+/K+ ATPase).
OS Artemia sanfranciscana (Brine shrimp) (Artemia franciscana).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Branchiopoda; Anostraca;
OC Artemiidae; Artemia.
OX NCBI_TaxID=6661;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92039032; PubMed=1657719;
RA Macias M.T., Martinez J.L., Palmero I., Sastre L.;
RT "Cloning of a cDNA encoding an Artemia franciscana Na/K ATPase alpha-
RT subunit.";
RL Gene 105:197-204(1991).
CC -!- FUNCTION: THIS IS THE CATALYTIC COMPONENT OF THE ACTIVE ENZYME,
CC WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE EXCHANGE OF
CC NA AND K IONS ACROSS THE PLASMA MEMBRANE. THIS ACTION CREATES THE
CC ELECTROCHEMICAL GRADIENT OF NA AND K, PROVIDING THE ENERGY FOR
CC ACTIVE TRANSPORT OF VARIOUS NUTRIENTS.

CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + Na(+) (In) + K(+) (Out) = ADP +
CC phosphate + Na(+) (Out) + K(+) (In).
CC -!- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
CC AND GAMMA.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IIC.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X56650; CAA39972.1; -.
CC PIR: JH0470; JH0470.
CC HSSP: P04191; IEUL.
CC InterPro: IPR001757; ATPase_E1-E2.
CC InterPro: IPR006069; Cation_ATPase.
CC InterPro: IPR006068; Cation_ATPase_C.
CC InterPro: IPR004034; Cation_ATPase_N.
CC InterPro: IPR005834; Hydrolase.
CC InterPro: IPR005775; Na/K_ATPase_alph.
CC Pfam: PF00690; Cation_ATPase_N; 1.
CC Pfam: PF00122; E1-E2_ATPase; 1.
CC Pfam: PF00702; Hydrolase; 1.
CC PRINTS: PR00119; CATATPASE.
CC PRINTS: PR00121; NAKATPASE.
CC TIGRFAMS: TIGR01106; ATPase_P-type; 1.
CC TIGRFAMS: TIGR01494; ATPase_P-type; 4.
CC PROSITE: PS00154; ATPase_E1-E2; 1.
KW Hydrolase; Sodium/potassium transport; Transmembrane;
KW Phosphorylation; ATP-binding.
FT TRANSMEM 76 96 BY SIMILARITY.
FT TRANSMEM 110 126 BY SIMILARITY.
FT TRANSMEM 272 294 BY SIMILARITY.
FT TRANSMEM 301 329 BY SIMILARITY.
FT TRANSMEM 768 791 BY SIMILARITY.
FT TRANSMEM 828 855 BY SIMILARITY.
FT TRANSMEM 897 918 BY SIMILARITY.
FT TRANSMEM 934 959 BY SIMILARITY.
FT MOD_RES 357 357 PHOSPHORYLATION (PROBABLE).
FT BINDING 489 489 ATP (BY SIMILARITY).
FT SEQUENCE 1004 AA; 110699 MW; CE4E6BCE19A78C7 CRC64;

Query Match 40.0%; Score 44; DB 1; Length 1004;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 VFGLKQDWGATIKD 20
   |||: |||: |
Db 864 LFGLRKHWDSDRAVND 878

RESULT 9
AROE_SYNY3
ID AROE_SYNY3          STANDARD;          PRT;          290 AA.
AC P74591;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Shikimate 5-dehydrogenase (EC 1.1.1.25).
GN AROE OR SLR1559.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
```



```

RESULT 13
ID YML6_MYCLE STANDARD; PRT; 307 AA.
AC 032960;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ML0860.
GN ML0860 OR MLCB32.18.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
EX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: BELONGS TO THE UPF0105 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; Z98741; CAB11383.1; -
DR EMBL; A1583920; CAC31241.1; -
DR PIR; T44893; T44893.
DR Leproma; ML0860; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 307 AA; 32853 MW; EEF5BA102455DA2 CRC64;
Query Match 38.2%; Score 42; DB 1; Length 307;
Best Local Similarity 70.0%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 9 LKQDWDGATI 18
Db 143 LCQDWDGATL 152
RESULT 14
ID QUEA_DEIRA STANDARD; PRT; 362 AA.
AC 09RU19;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sadenosylmethionine:tRNA ribosyltransferase-isomerase (EC 5.---.--)
DE (Queuosine biosynthesis protein queA).
GN QUEA OR DR1577.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;

```

```

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).
CC -1- FUNCTION: Synthesizes OQ from preO1 in a single S-
CC adenosylmethionine-requiring step. The ribosyl moiety of AdoMet is
CC transferred and isomerized to the epoxycyclopentane residue of OQ
CC (by similarity).
CC -1- PATHWAY: Queuosine biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE QUEA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE002001; AAF11138.1; -
DR PIR; F75379; F75379.
DR TIGR; DR1577; -
DR HAMAP; MF-00113; -; 1.
DR InterPro; IPR003699; Queuosine synth.
DR Pfam; PF02547; Queuosine synth; 1.
KW Queuosine biosynthesis; Transferase; Isomerase; Complete proteome.
SQ SEQUENCE 362 AA; 39561 MW; FD45EACAF59A3B2F CRC64;
Query Match 38.2%; Score 42; DB 1; Length 362;
Best Local Similarity 47.1%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
Qy 3 TTGVFLKQDWDGATIK 19
Db 271 TTVRTLESAMDGAAYR 287
RESULT 15
ID DNAA_STRCO STANDARD; PRT; 656 AA.
AC P27902; Q9KXX4;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chromosomal replication initiator protein dnAA.
GN DNAA OR SCO3879 OR SCH18.16C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=92250416; PubMed=1577691;
RA Calcutt M.J., Schmidt F.J.;
RT "Conserved gene arrangement in the origin region of the Streptomyces
RT coelicolor chromosome.";
RL J. Bacteriol. 174:3220-3226(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

```

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION
 CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;
 CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS
 CC (DNAA BOX): 5'-TATATC(C/A)(C/A)A-3'. DNAA BINDS TO ATP AND TO
 CC ACIDIC PHOSPHOLIPIDS.
 CC -1- SIMILARITY: Belongs to the dnaA family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF187159; AAA26734.1; -;
 CC EMBL; AL939118; CAD55464.1; -;
 CC PIR; A41870; A41870.
 CC HAMAP; MF_00377; -; 1.
 CC InterPro; IPR003593; AAA_ATPase.
 CC InterPro; IPR001957; Bac_DnaA.
 CC Pfam; PF00308; bac_dnaA; 1.
 CC PRINTS; PR00051; DNAA.
 CC SMART; SM00382; AAA; 1.
 CC TIGRFAMs; TIGR00362; DnaA; 1.
 CC PROSITE; PS01008; DNAA; 1.
 CC DNA replication; DNA-binding; ATP-binding; Complete proteome.
 KW NP_BIND 357 364
 FT SEQUENCE 656 AA; 73182 MW; 6C1D5C0193D3C92B CRC64;
 SQ
 Query Match 38.2%; Score 42; DB 1; Length 656;
 Best Local Similarity 36.8%; Pred. No. 41;
 Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
 QY 2 TTGTGVEGLKQWDGATIKD 20
 |:|:| |:|:|
 Db 563 TSTAIMGATADYFGLTVED 581
 RESULT 16
 ID ALA2_HUMAN STANDARD; PRT; 1020 AA.
 AC P50993; Q07059; Q9UQ25;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Sodium/potassium-transporting ATPase alpha-2 chain precursor
 DE (EC 3.6.3.9) (Sodium pump 2) (Na+/K+ ATPase 2).
 GN ATP1A2 OR KIAA0778.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90008924; PubMed=2477373;
 RA Shull M.M., Pugh D.G., Lingrel J.B.;
 RT "Characterization of the human Na,K-ATPase alpha 2 gene and
 RT identification of intragenic restriction fragment length
 RT polymorphisms.";
 RL J. Biol. Chem. 264:17532-17543(1989).
 RN [2]
 RP SEQUENCE OF 211-249 FROM N.A.
 RC TISSUE=Leukocyte;
 RX MEDLINE=87231946; PubMed=3035563;

RA Shull M.M., Lingrel J.B.;
 RT "Multiple genes encode the human Na+,K+-ATPase catalytic subunit.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4039-4043(1987).
 RN [3]
 RP SEQUENCE OF 251-442 FROM N.A.
 RC TISSUE=Brain, and Placenta;
 RX MEDLINE=87247232; PubMed=3036582;
 RA Sverdlov E.D., Monastyrskaya G.S., Broude N.E., Ushkaryov Y.A.,
 RA Allikmets R.L., Melkov A.M., Smirnov Y.V., Malyshev I.V.,
 RA Dulobova I.E., Petrukhin K.E., Grishin A.V., Kijatkina N.I.,
 RA Kostina M.B., Sverdlov V.E., Modyanov N.N., Ovchinnikov Y.A.;
 RT "The family of human Na+,K+-ATPase genes. No less than five genes
 RT and/or pseudogenes related to the alpha-subunit.";
 RL FEBS Lett. 217:275-278(1987).
 RN [4]
 RP SEQUENCE OF 1-4 FROM N.A.
 RX MEDLINE=89153603; PubMed=2537767;
 RA Sverdlov E.D., Bessarab D.A., Malyshev I.V., Petrukhin K.E.,
 RA Smirnov Y.V., Ushkaryov Y.A., Monastyrskaya G.S., Broude N.E.,
 RA Modyanov N.N.;
 RT "Family of human Na+,K+-ATPase genes. Structure of the putative
 RT regulatory region of the alpha+-gene.";
 RL FEBS Lett. 244:481-483(1989).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:277-286(1998).
 CC -1- FUNCTION: This is the catalytic component of the active enzyme,
 CC which catalyzes the hydrolysis of ATP coupled with the exchange of
 CC sodium and potassium ions across the plasma membrane. This action
 CC creates the electrochemical gradient of sodium and potassium,
 CC providing the energy for active transport of various nutrients.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + Na(+)(In) + K(+)(Out) = ADP +
 CC phosphate + Na(+)(Out) + K(+)(In).
 CC -1- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
 CC AND GAMMA.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J05096; AAA51797.1; -;
 CC EMBL; M16795; AAA51799.1; -;
 CC EMBL; M27578; AAA35575.1; -;
 CC EMBL; M27571; AAA35575.1; JOINED.
 CC EMBL; M27576; AAA35575.1; JOINED.
 CC EMBL; Y07494; CAA68793.1; ALT_SEQ.
 CC EMBL; AB018321; BAA34498.2; -;
 CC PIR; A34474; A34474.
 CC HSSP; P04191; 1EUL.
 CC Genew; HGNC:800; ATP1A2.
 CC MIM; 182340; -;
 CC InterPro; IPR001757; ATPase_E1-E2.
 CC InterPro; IPR006069; Cation_ATPase.
 CC InterPro; IPR006068; Cation_ATPase_C.
 CC InterPro; IPR004014; Cation_ATPase_N.
 CC InterPro; IPR005834; Hydrolase.
 CC InterPro; IPR005775; Na/K_ATPase_alph.
 CC Pfam; PF00689; Cation_ATPase_C; 1.
 CC Pfam; PF00690; Cation_ATPase_N; 1.

DR pfam: PF00122; EI-E2 ATPase; 1.
 DR pfam: PF00702; Hydrolase; 1.
 DR PRINTS: PR00119; CAVATPASE.
 DR PRINTS: PR00121; NAKATPASE.
 DR TIGRFAMS: TIGR01106; ATPase-TIC-X-K; 1.
 DR TIGRFAMS: TIGR01494; ATPase_EI_E2; 1.
 DR PROSITE: PS00154; ATPase_EI_E2; 5.
 KW Hydrolase; Sodium/potassium transport; Transmembrane; Phosphorylation;
 KW Magnesium; Metal-binding; ATP-binding; Multigene family.
 FT PROPEP 1 5
 FT CHAIN 1 1020
 FT
 FT DOMAIN 6 85
 FT TRANSMEM 86 106
 FT DOMAIN 107 129
 FT TRANSMEM 130 150
 FT DOMAIN 151 286
 FT TRANSMEM 287 306
 FT DOMAIN 307 318
 FT TRANSMEM 319 336
 FT DOMAIN 337 369
 FT TRANSMEM 370 789
 FT DOMAIN 790 799
 FT TRANSMEM 800 820
 FT DOMAIN 821 840
 FT TRANSMEM 841 863
 FT DOMAIN 864 915
 FT TRANSMEM 916 935
 FT DOMAIN 936 948
 FT TRANSMEM 949 967
 FT DOMAIN 968 982
 FT TRANSMEM 983 1003
 FT DOMAIN 1004 1020
 FT MOD_RES 374 374
 FT MOD_RES 940 940
 FT BINDING 80 82
 FT METAL 714 714
 FT METAL 718 718
 FT SEQUENCE 1020 AA; 112265 MW; 112265 MW; AFBDE8EA94FE84FC3 CRC64;
 Query Match 38.2%; Score 42; DB 1; Length 1020;
 Best Local Similarity 40.0%; Pred. No. 66;
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 OY 6 VFLGKQWDGATIKD 20
 DB 880 LLGRLDWDRTMND 894
 RESULT 17
 GCST_NEIMB STANDARD; PRT; 366 AA.
 AC Q9K0L8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable aminomethyltransferase (EC 2.1.2.10) (glycine cleavage system
 T protein).
 GN GCVT OR NMB0574.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Clecko A., Parksey D.S., Blair E., Citti H., Clark E.B.,
 Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
 Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 "Complete genome sequence of *Neisseria meningitidis* serogroup B strain
 MC58";
 RL Science 287:1809-1815(2000).
 CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of
 CC glycine (By similarity).
 CC -!- CATALYTIC ACTIVITY: (6S)-tetrahydrofolate + S-
 CC methylmethyldihydropyrimidin = (6R)-5,10-
 CC methylenetetrahydrofolate + NH(3) + dihydropyrimidin.
 CC -!- SUBUNIT: The glycine cleavage system is composed of four proteins:
 CC P, T, L, and G (By similarity).
 CC -!- SIMILARITY: Belongs to the gcvT family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE002413; AAF41002.1;
 DR PIR: A81183; A81183.
 DR TIGR: NMB0574;
 DR HAMAP: MF_00259; -; 1.
 DR InterPro: IPR006222; GCV.T.
 DR InterPro: IPR006223; GcvT.
 DR Pfam: PF01571; GCV.T; 1.
 DR TIGRFAMS: TIGR00528; gcvT; 1.
 KW Transferase; Aminotransferase; Complete proteome.
 SQ SEQUENCE 366 AA; 39740 MW; 566A7BCD21691D91 CRC64;
 Query Match 37.7%; Score 41.5; DB 1; Length 366;
 Best Local Similarity 44.8%; Pred. No. 26;
 Matches 13; Conservative 2; Mismatches 3; Indels 11; Gaps 2;
 OY 2 TTTGVF--GLKQ-----DWDGATIK 19
 DB 312 TTSGVSPSLKQSLAIARVPKDFDGTAK 340
 RESULT 18
 GCST_NEIMA STANDARD; PRT; 368 AA.
 AC Q9JVF2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable aminomethyltransferase (EC 2.1.2.10) (glycine cleavage system
 T protein).
 GN GCVT OR NMA0758.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z2451 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parthill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies K., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Barrett B.G., Barrall B.G.;
 "Complete DNA sequence of a serogroup A strain of *Neisseria*
 RT meningitidis 2299";
 RL Nature 404:502-506(2000).
 CC -!- FUNCTION: The glycine cleavage system catalyzes the degradation of
 CC glycine (By similarity).
 CC -!- CATALYTIC ACTIVITY: (6S)-tetrahydrofolate + S-


```
Query Match          37.3%; Score 41; DB 1; Length 395;
Best Local Similarity 58.3%; Pred. No. 35;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 8 GLKQWDGATIK 19
   ||| ||| |||
DB 57 GIKQVWDVATLQ 68

RESULT 21
HPFG_PORGI
ID HPFG_PORGI STANDARD; PRT; 684 AA.
AC Q9S3Q2; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chaperrone protein htpg (Heat shock protein htpg) (High temperature
DE protein G).
GN HTPG.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33277;
RX MEDLINE=20187468; PubMed=10722592;
RA Lopatin D.E., Combs A., Sweier D.G., Fenno J.C., Dharmija S.;
RT "Characterization of heat-inducible expression and cloning of HtpG
RL Infect. Immun. 68:1980-1987(2000).
CC -1- FUNCTION: Molecular chaperrone. Has ATPase activity (By
CC similarity).
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: AF176245; AAD51118.1; -.
DR HSSP; P07900; 1YER.
DR HAMAP; MF_00505; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR Pfam; PF02518; HATPase_c; 1.
DR PRINTS; PR00183; HSP90; 1.
DR SMART; SM00387; HATPase_c; 1.
DR PROSITE; PS00298; HSP90; FALSE_NEG.
KW Chaperrone; ATP-binding; Heat shock.
FT DOMAIN 1 329 A; SUBSTRATE-BINDING (BY SIMILARITY).
FT DOMAIN 330 548 B (BY SIMILARITY).
FT DOMAIN 549 684 C.
SQ SEQUENCE 684 AA; 78219 MW; 8384A999F32378B6 CRC64;

Query Match          37.3%; Score 41; DB 1; Length 684;
Best Local Similarity 40.0%; Pred. No. 63;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 VFGLKQDWDGATIKD 20
   ||| ||| |||
DB 202 IFGKQEWKDGKMQD 216

RESULT 22
VP4_ROTGA
ID VP4_ROTGA STANDARD; PRT; 749 AA.
AC Q04916; 1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4).
DE S4.
GN Rotavirus (group B / strain ADRV) (Adult diarrhea rotavirus).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=12705;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=9323240; PubMed=8386274;
RX Mackow E.R., Werner-Eckert R., Fay M.E., Tao H., Chen G.-M.;
RT "Identification and baculovirus expression of the VP4 protein of the
RT human group B rotavirus ADRV.";
RL J. Virol. 67:2730-2738(1993).
CC -1- SUBCELLULAR LOCATION: Outer capsid.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: M91434; AAA47338.1; -.
DR InterPro; IPR000416; Cap_VP4.
DR Pfam; PF00426; VP4; 1.
KW Cap protein; glycoprotein.
FT CARBOHYD 29 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 407 407 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 681 681 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 698 698 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 749 AA; 84362 MW; D123527DEAE0F21 CRC64;

Query Match          37.3%; Score 41; DB 1; Length 749;
Best Local Similarity 57.1%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 TGVFGLKQDWDGAT 17
   :| ||| |||
DB 285 SGKAGLQDWRPAT 298

RESULT 23
YCBS_ECOLI
ID YCBS_ECOLI STANDARD; PRT; 866 AA.
AC P75857.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical outer membrane usher protein ycbS precursor.
DE YCBS OR B0940.
GN Escherichia coli.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=k12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
```

RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Oshima K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155 (1996).
CC -1- FUNCTION: COULD BE INVOLVED IN THE EXPORT AND ASSEMBLY OF THE
CC PUTATIVE YCBQ FIMBRIAL SUBUNIT ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000196; AAC74026.1; -.
DR EMBL; D90732; BAA35695.1; -.
DR PIR; C64834; C64834.
DR EcoGene; EG13711; ycbS.
DR InterPro; IPR000015; Fimb_usher.
DR Pfam; PF00577; Usher; 1.
DR PROSITE; PS01151; FIMBRIAL_USHER; 1.
DR Hypothetical protein; Outer membrane; Transmembrane; Fimbria;
KW Transport; Signal; Complete proteome.
FT SIGNAL 1 35
FT CHAIN 36 866
FT HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN
FT YCBS.
SQ SEQUENCE 866 AA; 95241 MW; 0004DC5E9F1F5796 CRC64;

Query Match 37.3%; Score 41; DB 1; Length 866;
Best Local Similarity 63.6%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 GLKQDWDGATI 18
ID ANFL_CHICK STANDARD; PRT; 196 AA.
AC P79775;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein ANF-1 (GANF).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Kazanskaya O.V., Severtzova E.A., Barth K.A., Ermakova G.V.,
RA Lukyanov S.A., Benyumov A.O., Pannese M., Boncinelli E., Wilson S.W.,
RA Zaraisky A.G.;
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL [2]

RP PRELIMINARY SEQUENCE OF 8-196 FROM N.A.
RC TISSUE=Forebrain;
RX MEDLINE=98038973; PubMed=9373136;
RA Kazanskaya O.V., Severtzova E.A., Barth K.A., Ermakova G.V.,
RA Lukyanov S.A., Benyumov A.O., Pannese M., Boncinelli E., Wilson S.W.,
RA Zaraisky A.G.;
RT "Anf: a novel class of vertebrate homeobox genes expressed at the
anterior end of the main embryonic axis.";
RL Gene 200:25-34 (1997).
CC -1- FUNCTION: MAY BE INVOLVED IN THE EARLY PATTERNING OF THE MOST
CC ANTERIOR REGION OF THE MAIN EMBRYONIC BODY AXIS.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC THE EMBRYONIC BODY AXIS DURING GASTRULATION AND NEURULATION.
CC -1- SIMILARITY: BELONGS TO THE ANF HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U65436; AAC24899.1; -.
DR HSSP; P06601; 1FJL.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR007104; Paired_homeo.
DR Pfam; PF00046; homeobox; 1.
DR PRODOM; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 119 178
FT HOMEBOX.
SQ SEQUENCE 196 AA; 22801 MW; 1584272DE04AEF9F CRC64;

Query Match 36.8%; Score 40.5; DB 1; Length 196;
Best Local Similarity 38.5%; Pred. No. 20;
Matches 10; Conservative 3; Mismatches 6; Indels 7; Gaps 1;

QY 1 NTTTT-----GVFGLKQDWDGATIK 19
ID PURL_ZYMMO STANDARD; PRT; 734 AA.
AC Q9RE06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase II (EC 6.3.5.3) (FGAM
synthase II).
EN PUR-O.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;
OC Sphingomonadaceae; Zymomonas.
OX NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 31821 / ZMA / CP4;
RA Um H.W., Kang H.S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-
ribosyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-
(formamido)-N(1)-(5-phospho-D-ribosyl)acetamide + L-glutamate.
CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
CC -1- SUBUNIT: Heterodimer of two subunits, purQ and purl.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF213822; AAF23789.1; -;
DR HAMAP; MF 00420; -; 1.
DR InterPro; IPR000728; AIRS_related.
DR Pfam; PF00586; AIRS; 2.
DR Pfam; PF02769; AIRS.C; 2.
KW Purine biosynthesis; Ligase; ATP-binding.
FT NP_BIND 106 117
SQ SEQUENCE 734 AA; 77679 MW; B71635E0F66A166 CRC64;

Query Match 36.8%; Score 40.5; DB 1; Length 734;
Best Local Similarity 58.8%; Pred. No. 82;
Matches 10; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 3 TTGVEGLKQDW-DGATI 18
I I I I I I I I I
Db 551 TIGGVLGLQDWDRDSTTI 567

RESULT 26
ID DPVD_HUMAN STANDARD; PRT; 1025 AA.
AC Q12882; Q16694; Q16761;
DC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Dihydropyrimidine dehydrogenase [NADP+] precursor (EC 1.3.1.2) (DPD)
DE (DHPDase) (Dihydrouracil dehydrogenase) (Dihydrothymine
DE dehydrogenase).
GN DPVD.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RA MEDLINE=94365020; PubMed=8083224;
RA Yokota H., Fernandez-Salguero P., Furuya H., Lin K., McBride O.W.,
RA Podschun B., Schnackerz K.D., Gonzalez F.J.;
RT "cDNA cloning and chromosome mapping of human dihydropyrimidine
RT dehydrogenase, an enzyme associated with 5-fluorouracil toxicity and
RL congenital thymine uraciluria.";
RL J. Biol. Chem. 269:23192-23196(1994).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=97280676; PubMed=9135003;
RA Johnson M.R., Wang K., Tillmanns S., Albin N., Diasio R.B.;
RA "Structural organization of the human dihydropyrimidine dehydrogenase
RT gene";
RL Cancer Res...57:1660-1663(1997).
[3]
RN SEQUENCE OF 581-635 FROM N.A.
RP TISSUE=Liver;
RA MEDLINE=97047101; PubMed=8892022;
RA Kreken P., van Kullenburg A.B.P., Meinsma R., Smit G.P.A.,
RA Bakker H.D., de Abreu R.A., van Gennip A.H.;
RT "A point mutation in an invariant splice donor site leads to exon
RT skipping in two unrelated Dutch patients with dihydropyrimidine
RT dehydrogenase deficiency.";
RL J. Inherit. Metab. Dis. 19:645-654(1996).
[4]
RN CHARACTERIZATION, AND PARTIAL SEQUENCE.
RP TISSUE=Liver;
RX MEDLINE=92381021; PubMed=1512248;
RA Lu Z.-H., Zhang R., Diasio R.B.;

*Purification and characterization of dihydropyrimidine dehydrogenase
from human liver.";
J. Biol. Chem. 267:17102-17109(1992).
[5]
RN VARIANTS ARG-29; TRP-235 AND HIS-886.
RX MEDLINE=98102836; PubMed=9439663;
RA "Reken P., van Kullenburg A.B.P., Meinsma R., van Gennip A.H.;

*Dihydropyrimidine dehydrogenase (DPD) deficiency: identification and
expression of missense mutations C29R, R886H and R235W.";
Hum. Genet. 101:333-338(1997).
[6]
RN VARIANTS ARG-29; TRP-235 AND HIS-886.
RX MEDLINE=97411371; PubMed=9266349;
RA "Reken P., van Kullenburg A.B.P., Meinsma R., van Gennip A.H.;

*Identification of novel point mutations in the dihydropyrimidine
dehydrogenase gene.";
J. Inherit. Metab. Dis. 20:335-338(1997).

CC -!- FUNCTION: INVOLVED IN PYRIMIDINE BASE DEGRADATION. CATALYZES THE
CC REDUCTION OF URACIL AND THYMINE. ALSO INVOLVED THE DEGRADATION OF
CC THE CHEMOTHERAPEUTIC DRUG 5-FLUOROURACIL.

CC -!- CATALYTIC ACTIVITY: 5,6-dihydrouracil + NADP(+) = uracil + NADPH.

CC -!- COFACTOR: TWO EACH OF FAD AND FMN. ALSO CONTAINS TWO 4FE-4S
CC CLUSTERS. CONTAINS APPROXIMATELY 33 IRON ATOMS PER MOLECULE.

CC -!- PATHWAY: INITIAL AND RATE-LIMITING ENZYME IN THE 3-STEP PATHWAY OF
CC URACIL AND THYMINE CATABOLISM AND IN THE PATHWAY LEADING TO THE
CC FORMATION OF BETA-ALANINE.

CC -!- SUBUNIT: Homodimer.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- TISSUE SPECIFICITY: FOUND IN MOST TISSUES WITH GREATEST ACTIVITY
CC FOUND IN LIVER AND PERIPHERAL BLOOD MONONUCLEAR CELLS.

CC -!- DISEASE: DEFECTS IN DPVD ARE THE CAUSE OF HEREDITARY THYMINE-
CC URACILURIA (ALSO KNOWN AS FAMILIAL PYRIMIDINEMIA), A DISEASE

CC CHARACTERIZED BY PERSISTENT URINARY EXCRETION OF EXCESSIVE AMOUNTS
CC OF URACIL, THYMINE AND 5-HYDROXYMETHYLURACIL. PATIENTS SUFFERING

CC FROM THIS DISEASE SHOW A SEVERE REACTION TO THE ANTICANCER DRUG 5-
CC FLUOROURACIL. THIS REACTION INCLUDES STOMACHITIS, DIARRHEA, LOSS,

CC THROMBOCYTOPENIA, HAIR LOSS, DIARRHEA, FEVER, MARKED WEIGHT TO
CC CEREBELLAR ATAXIA, AND NEUROLOGIC SYMPTOMS, PROGRESSING TO

CC SEMICOMA.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL; U09178; AAA57474.1; -;
CC EMBL; U20938; AAB51366.1; -;
CC EMBL; X95670; CAA64973.1; -;
CC PIR; A54718; A54718

CC Genew; HGNC:3012; DPVD.

CC MIM; 274270; -; C:cytoplasm; NAS.

CC GO; GO:0005737; C:cytoplasm; NAS.

CC GO; GO:001711; F:dihydropyrimidine dehydrogenase activity; NAS.

CC GO; GO:0006214; F:thymidine catabolism; NAS.

CC GO; GO:0006212; F:uracil catabolism; NAS.

CC InterPro; IPR001450; 4Fe4S_ferredoxin.

CC InterPro; IPR000759; Adrnx_reductase.

CC InterPro; IPR001295; DHO_dhl.

CC InterPro; IPR005720; DHO_dhl.

CC InterPro; IPR001327; FAD_pyr_redox.

CC InterPro; IPR003009; FMN_enzyme.

CC InterPro; IPR000103; Pyridine_redox_2.

CC Pfam; PF01180; DHodehase; 1.

CC Pfam; PF00037; fer4; 2.

CC PRINTS; PR00419; ADXRDTASE.

CC PRINTS; PR00368; FADPNR.

CC PRINTS; PR00469; PNDRTASEII.

CC TIGRFAMs; TIGR01037; pyrd_subl_fam; 1.

CC PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.

CC Oxidoreductase; NADP; Flavoprotein; FAD; FMN; Iron-sulfur; 4Fe-4S;

KW


```

FT CARBOHYD 78 78 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 191 AA; 21415 MW; 117387F7B94B9EE3 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 191;
Best Local Similarity 56.2%; Pred. No. 23;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 TTGFGGLKQDWDGATI 18
   ||| ||| : ||:
Db 133 TTGKTGLGQDFHMYATL 148

RESULT 30
BIOB_HELPJ
ID BIOB_HELPJ STANDARD; PRT; 282 AA.
AC Q9ZJK8:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).
GN BIOB OR JHP1298.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummano P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Malls S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -1- PATHWAY: Biotin biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE001553; AAD06876.1; -.
CC PIR; H71823; H71823.
CC InterPro; IPR002684; Biotin_synth.
CC InterPro; IPR006638; Elp3.
CC Pfam; PF04055; Radical_SAM; 1.
CC SMART; SM00729; Elp3; 1.
CC TIGRFAMS; TIGR00433; bioB; 1.
CC Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
CC FT METAL 17 17 IRON-SULFUR (POTENTIAL).
CC FT METAL 21 21 IRON-SULFUR (POTENTIAL).
CC FT METAL 24 24 IRON-SULFUR (POTENTIAL).
CC SEQUENCE 282 AA; 31490 MW; 3A07177B65AFAF1B CRC64;

Query Match 36.4%; Score 40; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVFGGLKQDWD 14
   ||| ||| : ||:
Db 157 GIFGLNESWE 166

RESULT 31
BIOB_HELPJ
ID BIOB_HELPJ STANDARD; PRT; 282 AA.
AC Q9ZJK8:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin synthase (EC 2.8.1.6) (Biotin synthetase).
GN BIOB OR HPI1406.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=99252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- CATALYTIC ACTIVITY: Dethiobiotin + sulfur = biotin.
CC -1- PATHWAY: Biotin biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE BIOTIN AND LIPOIC ACID SYNTHETASES
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE000640; AAD08448.1; -.
CC PIR; F64695; F64695.
CC TIGR; HPI1406; -.
CC InterPro; IPR002684; Biotin_synth.
CC InterPro; IPR006638; Elp3.
CC Pfam; PF04055; Radical_SAM; 1.
CC SMART; SM00729; Elp3; 1.
CC TIGRFAMS; TIGR00433; bioB; 1.
CC Biotin biosynthesis; Iron-sulfur; Transferase; Complete proteome.
CC FT METAL 17 17 IRON-SULFUR (POTENTIAL).
CC FT METAL 21 21 IRON-SULFUR (POTENTIAL).
CC FT METAL 24 24 IRON-SULFUR (POTENTIAL).
CC SEQUENCE 282 AA; 31475 MW; 8EAD8B801DF151EE CRC64;

Query Match 36.4%; Score 40; DB 1; Length 282;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 GVFGGLKQDWD 14
   ||| ||| : ||:
Db 157 GIFGLNESWE 166

RESULT 32
BIOB_HELPJ
ID BIOB_HELPJ STANDARD; PRT; 341 AA.
AC P75302;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG335.2 homolog (P01_orf341).
GN MPN483 OR MP359.
OS Mycoplasma pneumoniae.

```

CC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelfreich R., Hilbert H., Piagens H., Pirkel E., Li B.-C.,
 RA Herrmann R.;
 RA "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC STRONG, TO M.GENITALIUM WC35.2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000034; AAB96007.1;
 DR PIR; S73685; S73685.
 DR InterPro; IPR001173; Glyco_trans.2.
 DR Pfam; PF00535; Glycos.transf.2.1.
 DR Hypothetical protein; Transferase; Glycosyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 341 AA; 40414 MW; C209F50D714CB3D0 CRC64;

Query Match 36.4%; Score 40; DB 1; Length 341;
 Best Local Similarity 60.0%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 9 LKODWDGATV 18
 |||||:
 Db 26 LKODWNGVKV 35

RESULT 33
 UBPE_CAEEL
 ID URPE_CAEEL STANDARD; PRT; 489 AA.
 AC 017361; O45248;
 DT 01-NOV-1999 (Rel. 35, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 14 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 14) (Ubiquitin-specific processing protease 14)
 DE (Deubiquitinating enzyme 14).
 GN TGT-1 OR CL3B4.2
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Deshpande K.L., Katze J.R.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Percy C., Smye R.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC -!- CAUTION: Was originally (Ref.1) thought to be a queuine tRNA-
 CC ribosyltransferase.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32223; AAA74956.1;
 DR EMBL; Z83236; CAB05785.1;
 DR EMBL; Z81468; CAB05785.1; JOINED.
 DR EMBL; Z81468; CAB03876.1;
 DR EMBL; Z83236; CAB03876.1; JOINED.
 DR PIR; T19227; T19227.
 DR Knappep; C13B4.2; CE15615.
 DR MPROFS; C19.UFW;
 DR InterPro; IPR000626; Ubiquitin.
 DR InterPro; IPR001394; UCH-2.
 DR Pfam; PF00440; ubiquitin; 1.
 DR Pfam; PF00443; UCH; 1.
 DR SMART; SM0213; UHQ; 1.
 DR PROSITE; PS00372; UCH_2.1; 1.
 DR PROSITE; PS00373; UCH_2.2; 1.
 DR PROSITE; PS50235; UCH_2.3; 1.
 KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 111
 FT ACT_SITE 400
 FT ACT_SITE 409
 FT CONFLICT 398
 FT CONFLICT 398
 SQ SEQUENCE 489 AA; 55877 MW; 4433F2CC59703A9B CRC64;

Query Match 36.4%; Score 40; DB 1; Length 489;
 Best Local Similarity 66.7%; Pred. No. 64;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 12 DWGDGATKD 20
 |||||:
 Db 54 DWEGITIKE 62

RESULT 34
 AMPA_BACSU
 ID AMPA_BACSU STANDARD; PRT; 500 AA.
 AC 032106;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable cytosol aminopeptidase (EC 3.4.11.1) (Leucine aminopeptidase)
 DE (LAP) (Leucyl aminopeptidase).
 GN PEPA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignonell C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V.C., Carlier N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson F.I.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritze C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.I., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Hilbert H., Holsappel S., Hosono S., Hulech J., Harwood C.R., Henaut A.,
 RA Jorjabad H., Karamata D., Kasahara Y., Kjaer-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Meliadi R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Pressecan E., Fujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Tannier F., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Verstra P., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*,"
 RL Nature 390:249-256(1997).
 CC -!- FUNCTION: PRESUMABLY INVOLVED IN THE PROCESSING AND REGULAR
 CC TURNOVER OF INTRACELLULAR PROTEINS. CATALYZES THE REMOVAL OF
 CC UNSUBSTITUTED N-TERMINAL AMINO ACIDS FROM VARIOUS PEPTIDES (BY
 CC SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, xaa-|-
 CC xbb-, in which xaa is preferably Leu, but may be other amino acids
 CC including pro although not Arg or Lys, and xbb may be pro.
 CC -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M17.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z99120; CAB15195.1; -;
 DR PIR; F70012; F70012.
 DR HSP; P00727; 1LAM.
 DR MEROPS; M17; UPW; -;
 DR Subtilist; BG13970; pepA.
 DR HAMAP; MF_00181; -; 1.
 DR InterPro; IPR000819; Peptidase_M17.
 DR Pfam; PF00883; Peptidase_M17; 1.
 DR Pfam; PF02789; Peptidase_M17_N; 1.
 DR PRINTS; PR00481; LAMNOPPTDASE.
 DR PROSITE; PS00631; CYTOSOL_AP; 1.
 KW Hydroxylase; Aminopeptidase; Manganese; Complete proteome.
 FT METAL 261 261 MANGANESE 2 (BY SIMILARITY).
 FT METAL 266 266 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 284 284 MANGANESE 2 (BY SIMILARITY).
 FT METAL 343 343 MANGANESE 1 (BY SIMILARITY).
 FT METAL 345 345 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT ACT_SITE 273 273 POTENTIAL.
 FT ACT_SITE 347 347 POTENTIAL.
 SQ SEQUENCE 500 AA; 53657 MW; 3E82968F6656559 CRC64;
 Query Match 36.4%; Score 40; DB 1; Length 500;
 Best Local Similarity 46.7%; Pred. No. 65;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 2 TTGTGVLGKQDWDGA 16
 | : : | : | : |
 Db 274 TKSGIVGKMSDMDGA 288

RESULT 35

GUAA_CORGL
 ID GUAA_CORGL STANDARD; PRT; 523 AA.
 AC Q8NSR1;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (Glutamine
 DE amidotransferase) (GMP synthetase).
 GN GUAA OR CGL0607.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 RX NCBI_TaxID=1718;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
 RA Nakagawa S.;
 RT "Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032,"
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ATP + xanthosine 5'-phosphate + L-glutamine +
 CC H(2)O -> AMP + diphosphate + GMP + L-glutamate.
 CC -!- PATHWAY: GMP biosynthesis.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMP SYNTHASE
 CC FAMILY.
 CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AP005275; BAB98000.1; -;
 DR HAMAP; MF_00344; -; 1.
 DR InterPro; IPR006220; Anth_synthII.
 DR InterPro; IPR001317; CPS_GATase.
 DR InterPro; IPR000991; GATase_1.
 DR InterPro; IPR001674; GMP_synth_C.
 DR InterPro; IPR004739; GMPsynthase_N.
 DR Pfam; PF00117; GATase; 1.
 DR Pfam; PF00958; GMP_synth_C; 1.
 DR PRINTS; PR00097; ANTSNTHASEII.
 DR PRINTS; PR00099; CPSGATASE.
 DR PRINTS; PR00096; GATASE.
 DR TIGRFAMS; TIGR00884; guaA_Cterm; 1.
 DR TIGRFAMS; TIGR00888; guaA_Nterm; 1.
 DR PROSITE; PS00442; GATASE_TypeI; 1.
 KW Ligase; GMP biosynthesis; Purine biosynthesis; ATP-binding;
 KW Glutamine amidotransferase; Complete proteome.
 FT DOMAIN 1 197 GLUTAMINE AMIDOTRANSFERASE.
 FT NP_BIND 227 233 ATP (BY SIMILARITY).
 FT DOMAIN 231 398 GMP-BINDING (BY SIMILARITY).
 FT ACT_SITE 86 86 GATASE (BY SIMILARITY).
 FT ACT_SITE 172 172 GATASE (BY SIMILARITY).
 FT ACT_SITE 174 174 GATASE (BY SIMILARITY).
 SQ SEQUENCE 523 AA; 55986 MW; 7B8E580D0CDE0A58 CRC64;
 Query Match 36.4%; Score 40; DB 1; Length 523;
 Best Local Similarity 41.2%; Pred. No. 68;
 Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;
 QY 4 TCVFGLKQDWDGATIKD 20
 | : | : | : | : | : |
 Db 190 TEIAGLEQNWTAAIAE 206

Search completed: October 6, 2003, 07:44:01
 Job time : 10.6716 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:03 ; Search time 35.3731 Seconds
(without alignments)
89.744 Million cell updates/sec

Title: US-09-765-739A-1

Perfect score: 110

Sequence: 1 KSTGVGFLKHDWDGSPILK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

A_Geneseq_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	110	100.0	20	ABG30743	Ehrlichia canis pe
2	110	100.0	278	AA171477	Ehrlichia canis im
3	110	100.0	278	AAU96100	Ehrlichia canis p2
4	110	100.0	307	AAU06961	E. canis p30-1 pro
5	110	100.0	307	ABG77952	Ehrlichia canis ou
6	79	71.8	280	AAW51094	Ehrlichia chaffeen
7	79	71.8	280	AAU06962	E. canis p30-2 pro
8	79	71.8	280	AAAB36188	Ehrlichia chaffeen
9	79	71.8	280	AA171479	Ehrlichia canis im

10	79	71.8	280	22	AAU04198	Variable surface a
11	79	71.8	280	23	ABG77953	Ehrlichia canis p2
12	79	71.8	280	23	AAU96102	Ehrlichia canis p2
13	79	71.8	280	23	AAU73417	Ehrlichia chaffeen
14	76	69.1	280	20	AAU06948	E. chafeensis OMP-
15	76	69.1	280	23	ABG77940	Ehrlichia chaffeen
16	76	69.1	280	23	AAU96110	Ehrlichia chaffeen
17	73	66.4	19	23	ABG30749	Ehrlichia chaffeen
18	72	65.5	20	23	ABG30745	Ehrlichia chaffecu
19	72	65.5	256	20	AAU06942	E. chafeensis p28
20	72	65.5	256	23	ABG77966	protein encoded by
21	72	65.5	276	19	AAW51095	Ehrlichia chaffeen
22	72	65.5	276	21	AAAB36189	Ehrlichia chaffeen
23	72	65.5	276	22	AAU04199	Variable surface a
24	72	65.5	280	19	AAW51089	Ehrlichia chaffeen
25	72	65.5	280	21	AAAB36183	Ehrlichia chaffeen
26	72	65.5	280	22	AAU04193	Major antigenic pr
27	72	65.5	281	20	AAU06943	E. chafeensis OMP-
28	72	65.5	281	23	ABG77935	Ehrlichia chaffeen
29	72	65.5	281	23	AAU96105	Ehrlichia chaffeen
30	72	65.5	281	23	AAU73418	Ehrlichia chaffeen
31	70	63.6	20	23	ABG30744	Ehrlichia canis pe
32	70	63.6	288	20	AAU06959	E. canis p30 prote
33	70	63.6	288	23	ABG77950	Ehrlichia canis ou
34	60	54.5	287	19	AAW51088	Cowdria ruminantiu
35	60	54.5	287	21	AAAB36182	Cowdria ruminantiu
36	60	54.5	287	22	AAU04192	Major antigenic pr
37	58	52.7	19	23	ABG30747	Ehrlichia chaffeen
38	58	52.7	284	23	AAU96111	Cowdria ruminantiu
39	58	52.7	286	19	AAW51092	Ehrlichia chaffeen
40	58	52.7	286	20	AAU06946	E. chafeensis OMP-
41	58	52.7	286	21	AAAB36186	Ehrlichia chaffeen
42	58	52.7	286	22	AAU04196	Variable surface a
43	58	52.7	286	23	ABG77938	Ehrlichia chaffeen
44	58	52.7	286	23	AAU96108	Ehrlichia chafeens
45	57	51.8	276	20	AAU73415	Ehrlichia chaffeen
46	57	51.8	276	23	ABG77955	E. canis p30-4 pro
47	57	51.8	276	23	AAU96117	Ehrlichia canis ou
48	57	51.8	276	23	AAU96117	Ehrlichia canis p2
49	53	48.2	18	23	ABG30748	Ehrlichia chaffeen
50	53	48.2	278	19	AAW51093	Ehrlichia chaffeen
51	53	48.2	278	20	AAU06947	E. chafeensis OMP-
52	53	48.2	278	21	AAAB36187	Ehrlichia chaffeen
53	53	48.2	278	22	AAU04197	Variable surface a
54	53	48.2	278	23	AAE31090	Ehrlichia ruminant
55	53	48.2	278	23	ABG77939	Ehrlichia chaffeen
56	53	48.2	278	23	AAU96109	Ehrlichia chafeens
57	53	48.2	278	23	AAU73416	Ehrlichia chaffeen
58	52	47.3	18	23	ABG30746	Ehrlichia chaffeen
59	52	47.3	280	20	AAU06945	E. chafeensis OMP-
60	52	47.3	280	23	ABG77937	Ehrlichia chaffeen
61	52	47.3	280	23	AAU96107	Ehrlichia chaffeen
62	52	47.3	280	23	AAU73414	Ehrlichia chaffeen
63	50	45.5	904	20	AAU21976	Senescence-associa
64	48	43.6	132	22	AAU04201	Variable surface a
65	48	43.6	133	19	AAW51097	Ehrlichia canis VS
66	48	43.6	133	21	AAAB36191	Ehrlichia canis pa
67	48	43.6	133	21	AAU71480	Ehrlichia canis im
68	48	43.6	133	23	AAU96103	Ehrlichia canis p2
69	48	43.6	283	21	AAU71478	Ehrlichia canis im
70	48	43.6	283	23	AAU96101	Ehrlichia canis p2
71	46	41.8	59	22	ABG49936	Human liver peptid
72	46	41.8	59	22	ABG29909	Peptide #2560 enco
73	46	41.8	59	22	ABG35087	Peptide #2593 enco
74	46	41.8	59	22	ABG20505	Protein #2504 enco
75	46	41.8	59	22	AAW55908	Human brain expres
76	46	41.8	59	22	AAW68278	Human bone marrow
77	46	41.8	59	22	AAU16101	Peptide #2535 enco
78	46	41.8	59	22	AAW28592	Peptide #2629 enco
79	46	41.8	59	22	AAW03828	Peptide #2510 enco
80	46	41.8	59	23	ABG37819	Human peptid enco
81	46	41.8	65	22	AAAB36188	Propionibacterium
82	46	41.8	850	17	AAAR91307	Helicobacter pylor

83 46 41.8 859 15 AAR53268
 84 46 41.8 859 16 AAR72594
 85 46 41.8 1181 15 AAR53269
 86 46 41.8 1181 16 AAR72593
 87 46 41.8 1181 17 AAR91307
 88 45.5 41.4 56 22 AAM91384
 89 45 40.9 15 23 ABB83444
 90 45 40.9 132 22 AAU33903
 91 45 40.9 132 22 AAU36710
 92 45 40.9 132 22 AAU37282
 93 45 40.9 132 22 AAU37558
 94 45 40.9 199 22 AAM38783
 95 45 40.9 199 23 ABB83443
 96 45 40.9 207 22 AAM25907
 97 45 40.9 230 22 AAM0569
 98 45 40.9 452 23 ABB89979
 99 45 40.9 562 20 AAY30895
 100 45 40.9 570 15 AAR54219

ALIGNMENTS

RESULT 1

ABG30743
 ID ABG30743 standard; Peptide; 20 AA.

XX AC ABG30743;

XX DT 21-OCT-2002 (first entry)

XX DE Ehrlichia canis peptide fragment #1.

XX KW Antibody detection; monoclonal antibody; polyclonal antibody.

XX OS Ehrlichia canis.

XX PN WO200257794-A2.

XX PD 25-JUL-2002.

XX PF 16-JAN-2002; 2002WO-US01395.

XX PR 18-JAN-2001; 2001US-0765739.

XX PA (INDEX) IDEXX LAB INC.

XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;

XX DR WPI; 2002-599730/64.

XX PS New composition of matter comprising a polypeptide, useful in detecting the presence of antibodies to Ehrlichia canis or chaffensis, or in detecting or quantifying the presence of Ehrlichia infection in mammals

XX Claim 1; Page 5; 29pp; English.

XX The invention relates to a composition of matter comprising a polypeptide isolated from Ehrlichia species. The composition can be used for detecting the presence of antibodies to Ehrlichia, comprising contacting one or more polypeptides with a test sample suspected of comprising antibodies to Ehrlichia, under conditions that allow polypeptide/antibody complexes to form and detecting the complexes, where the detection of polypeptide/antibody complexes is an indication that antibodies to Ehrlichia are present in the test sample. The composition is useful for detecting or quantifying the presence of E. canis or E. chaffeensis infection in mammals. The polypeptides can be used to develop monoclonal and/or polyclonal antibodies that can be employed in assay systems and in the generation of chimeric antibodies for therapeutic use or other similar applications. This sequence represents an E. canis peptide fragment used in the composition of the invention.

SQ Sequence 20 AA;

Query Match 100.0%; Score 110; DB 23; Length 20;
 Best Local Similarity 100.0%; Pred. No. 2.5e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTGVVFLKHDWDGSPILK 20
 |||||
 DB 1 KSTGVVFLKHDWDGSPILK 20

RESULT 2

AAY71477

ID AAY71477 standard; Protein; 278 AA.

XX AC AAY71477;

XX DT 12-OCT-2000 (first entry)

XX DE Ehrlichia canis immunoreactive protein Eca28-1.

XX KW Homologous 28-kDa protein gene; Eca28-1; immunoreactive; vaccine;
 p28 gene; polymorphic multiple gene family; immunoprotective antigen;
 antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
 tick-borne rickettsial disease; serodiagnosis.

XX OS Ehrlichia canis.

XX FH Key Location/Qualifiers

FT Peptide 1..23

FT Protein /label= Signal_peptide

FT Cleavage-site 24..278

FT Cleavage-site 21..23 /label= Mature_Eca28-1_28-kDa_protein

FT Cleavage-site 23..25 /label= Signal_peptidase_cleavage_site

FT /note= "Additional putative signal peptidase cleavage site"

FT Region 27..34 /label= Variable_region_1

FT /note= "Contains immunoreactive peptides"

FT Region 76..84 /label= Variable_region_2

FT /note= "Contains immunoreactive peptides"

FT Region 148..156 /label= Variable_region_3

FT /note= "Contains immunoreactive peptides"

FT Region 246..258 /label= Variable_region_4

FT /note= "Contains immunoreactive peptides"

XX WO200032745-A2.

XX PD 08-JUN-2000.

XX PF 24-NOV-1999; 99WO-US28075.

XX PR 30-NOV-1998; 98US-0201458.

XX PR 03-MAR-1999; 99US-0261358.

XX PA (RERE-) RES DEV FOUND.

XX PI Walker DH, Yu X, McBride JW;

XX DR WPI; 2000-412298/35.

XX DR N-PSDB; AAD01292.

XX PT Ehrlichia canis antigens useful for vaccinating against canine ehrlichiosis in dogs -

XX PS Claim 12; Fig 1; 86pp; English.

XX CC The patent relates to homologous 28-kilodalton (kDa) protein genes of

CC Ehrlichia canis, designated ECA28SA1, ECA28SA2, ECA28SA3, ECA28-1 and
CC ECA28-2. These genes are members of a polymorphic multiple gene family
CC and contained in a single locus of 5.592 kb. The 28-kDa proteins are
CC immunoreactive with anti-E. canis serum hence are important
CC immunoprotective antigens. The protein is useful for vaccinating
CC against E. canis infections such as canine ehrlichiosis in dogs.
CC Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
CC tick-borne rickettsial disease of dogs. ECA28-1 is conserved amongst
CC different strains of E. canis and hence useful for serodiagnosis of
CC canine ehrlichiosis. The present sequence is a E. canis
CC ECA28-1 30-kDa protein which is post-translationally modified to a
CC mature 28-kDa protein by cleavage of N-terminal signal sequence.
XX
SQ Sequence 278 AA;
Query Match 100.0%; Score 110; DB 21; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTVGVLGLKHDWDGSPILK 20
|||||
Db 60 KSTVGVLGLKHDWDGSPILK 79
RESULT 3
AAU96100
ID AAU96100 standard; Protein; 278 AA.
AC AAU96100;
XX
DT 02-JUL-2002 (first entry)
XX
DE Ehrlichia canis p28-7.
XX
KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
KW antibacterial.
XX
OS Ehrlichia canis.
XX
PN WO200222782-A2.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28759.
XX
PR 12-SEP-2000; 2000US-0660587.
XX
PA (RERE-) RES DEV FOUND.
XX
PI Walker DH, Yu X, McBride JW;
XX
XX WPI: 2002-351882/38.
DR N-PSDB; ABK68852.
XX
PT New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections
XX
PS Claim 16; Figure 1; 106pp; English.
XX
CC The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
SQ Sequence 278 AA;
Query Match 100.0%; Score 110; DB 23; Length 278;
Best Local Similarity 100.0%; Pred. No. 5.3e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTVGVLGLKHDWDGSPILK 20
|||||
Db 60 KSTVGVLGLKHDWDGSPILK 79
RESULT 4
AAU06961
ID AAU06961 standard; Protein; 307 AA.
AC AAU06961;
XX
DT 05-JUL-1999 (first entry)
XX
DE E. canis P30-1 protein.
XX
KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia canis.
XX
PN WO9913720-A1.
XX
PD 25-MAR-1999.
XX
PF 18-SEP-1998; 98WO-US19600.
XX
PR 19-SEP-1997; 97US-0059353.
XX
PA (OHIS) UNIV OHIO STATE.
XX
PI Ohashi N, Rikihisa Y;
XX
DR WPI: 1999-254290/21.
DR N-PSDB; AAX34761.
XX
PT Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
PS Disclosure; Fig 21B; 55pp; English.
XX
CC The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAU06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAU06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 307 AA;
Query Match 100.0%; Score 110; DB 20; Length 307;
Best Local Similarity 100.0%; Pred. No. 5.9e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTVGVLGLKHDWDGSPILK 20
|||||
Db 89 KSTVGVLGLKHDWDGSPILK 108
RESULT 5
ABG77952
ID ABG77952 standard; Protein; 307 AA.
XX
AC ABG77952;
XX
DT 15-NOV-2002 (first entry)
XX
DE Ehrlichia canis outer membrane protein (P30F) #3.
XX
KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
OS Ehrlichia canis.
XX
PN US2002120115-A1.

XX 29-AUG-2002.
 XX 28-JAN-2002; 2002US-0059964.
 XX 19-MAY-1999; 99US-0314701.
 XX (RIKI/) RIKIHISA Y.
 XX (OHAS/) OHASHI N.
 XX Rikihisa Y, Ohashi N;
 XX WPI; 2002-618954/66.
 XX N-PSDB; ABS63293.
 XX Isolated polynucleotide encoding an outer membrane protein of E.canis
 XX or E.chaffeensis used in the diagnosis of infection -
 XX Claim 10; Fig 21B; 49pp; English.
 XX The invention relates to an isolated polynucleotide encoding an outer
 XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 XX in the diagnosis of infection. An infection such as human ehrlichiosis or
 XX canine ehrlichiosis can be diagnosed by providing a serum sample from the
 XX patient, providing a polypeptide or mixture of polypeptides, contacting
 XX the sample with the polypeptide and assaying for the formation of a
 XX complex between antibodies in the serum sample and the polypeptide, where
 XX formation of a complex is indicative of infection with E. chaffeensis.
 XX This sequence represents an Ehrlichia outer membrane protein of the
 XX invention.
 XX SQ Sequence 307 AA;
 Query Match 100.0%; Score 110; DB 23; Length 307;
 Best Local Similarity 100.0%; Pred No. 5.9e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KSTVGVLKHDWDGSPILK 20
 DB 89 KSTVGVLKHDWDGSPILK 108
 RESULT 6
 AAW51094
 ID AAW51094 standard; Protein; 280 AA.
 XX AC AAW51094;
 XX DT 14-SEP-1998 (first entry)
 XX DE Ehrlichia chaffeensis VSA4 protein.
 XX KW MAP1 homologue; variable surface antigen; VSA4; rickettsia;
 XX KW DNA vaccine.
 XX OS Ehrlichia chaffeensis.
 XX PF Key Location/Qualifiers
 XX FH Peptide 1..25
 XX FT /note- "putative signal peptide"
 XX PN W09816554-A1.
 XX PD 23-APR-1998.
 XX PF 17-OCT-1997; 97WO-US19044.
 XX PR 17-OCT-1996; 96US-0733230.
 XX PA (UYFL) UNIV FLORIDA.
 XX PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 XX PI Nyika A, Rurangirwa FR;

XX WPI; 1998-251232/22.
 XX N-PSDB; AAV07179.
 XX Composition containing nucleic acid encoding rickettsial antigen -
 XX useful for, e.g. stimulating protective immune response in humans or
 XX animals
 XX Claim 3; Fig 2B; 39pp; English.
 XX This is the full-length variable surface antigen VSA4 protein of
 XX Ehrlichia chaffeensis. Its amino acid sequence was deduced from a
 XX partial open reading frame (ORF4) of a genomic locus (see AAV07179)
 XX of E. chaffeensis that was obtained on the basis of homology to the
 XX major antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium.
 XX This genomic locus included 5 ORFs encoding similar, but
 XX non-identical proteins (see AAW51091-95). A claimed composition
 XX comprises a nucleic acid (see AAV07176-82) encoding a polypeptide
 XX (see AAW51088-99) that elicits a protective immune response against a
 XX rickettsial pathogen. The nucleic acid is used, in human or
 XX veterinary medicine, in vaccines to protect against Rickettsia,
 XX Ehrlichia, Anaplasma and Cowdria species. The Ehrlichia antigenic
 XX polypeptides can be used diagnostically to detect antibodies
 XX associated with Ehrlichia infection (claimed).
 XX SQ Sequence 280 AA;
 Query Match 71.8%; Score 79; DB 19; Length 280;
 Best Local Similarity 73.7%; Pred No. 7.8e-05;
 Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 STVGVLKHDWDGSPILK 20
 DB 61 TTIGVGLKQDWDGSPILK 79
 RESULT 7
 AAY06962
 ID AAY06962 standard; Protein; 280 AA.
 XX AC AAY06962;
 XX DT 05-JUL-1999 (first entry)
 XX DE E. canis P30-2 protein.
 XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 XX KW detection; dog.
 XX OS Ehrlichia canis.
 XX PN W09913720-A1.
 XX PD 25-MAR-1999.
 XX PF 18-SEP-1998; 98WO-US19600.
 XX PR 19-SEP-1997; 97US-0059353.
 XX PA (OHTS) UNIV OHIO STATE.
 XX PI Ohashi N, Rikihisa Y;
 XX WPI; 1999-254290/21.
 XX N-PSDB; AAX34762.
 XX Novel outer membrane proteins from Ehrlichia chaffeensis and
 XX Ehrlichia canis
 XX Disclosure; Fig 22B; 55pp; English.
 XX The invention provides isolated outer membrane proteins (OMP) from
 XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part

CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the p30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
SQ Sequence 280 AA;

Query Match 71.8%; Score 79; DB 20; Length 280;
Best Local Similarity 82.4%; Pred. No. 7.8e-05;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
DB 60 STVGVEGLKHDWNGGT1 76

RESULT 8
AAB36188
ID AAB36188 standard; Protein; 280 AA.

AC AAB36188;
DT 02-MAR-2001 (first entry)
DE Ehrlichia chaffeensis partial VSA4.

XX
KW Ehrlichia chaffeensis; VSA4; variable surface antigen 4; MAP1;
KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
KW Rickettsia; Cowdria; Anaplasma; map2; lhworf3; 4hworf1; 18hworf1;
KW 3gdorf3.

OS Ehrlichia chaffeensis.

XX WO200065063-A2.

PN 02-NOV-2000.

XX 21-APR-2000; 2000WO-US10886.

XX 22-APR-1999; 99US-0130725.

XX (UYFL) UNIV FLORIDA.

XX Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
XX Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;

XX WPI: 2000-679675/66.
XX N-PSDB; AAC68705.

XX New polynucleotides useful as DNA vaccines for conferring immunity to
XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium,
XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

Claim 3; Page 45-46; 63pp; English.

XX The present sequence shows a high degree of similarity to the major
XX antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
XX used in a vaccines to protect animals or humans against rickettsial
XX diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
XX Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
XX protective against the rickettsial pathogen. The nucleic acid vaccines
XX can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
XX Cowdria ruminantium genes designated map 2, lhworf3, 4hworf1, 18hworf1
XX and 3gdorf3 may be used in therapeutic and diagnostic applications. The
XX polypeptides are useful for detecting antibodies associated with
XX infection by a rickettsial pathogen whilst the polynucleotides may be
XX used to detect the presence of rickettsial nucleic acids.

XX Sequence 280 AA;

Query Match 71.8%; Score 79; DB 21; Length 280;
Best Local Similarity 73.7%; Pred. No. 7.8e-05;

Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 2 STVGVEGLKHDWDGSP1X 20
DB 61 TTIGVFGKQDWDGSGTISK 79

RESULT 9
AAY71479
ID AAY71479 standard; Protein; 280 AA.

XX AAY71479;

XX 12-OCT-2000 (first entry)

XX Ehrlichia canis immunoreactive protein Eca28SA3.

XX Homologous 28-kDa protein gene; Eca28SA3; immunoreactive; vaccine;
XX p28 gene; polymorphic multiple gene family; immunoprotective antigen;
XX antibacterial; canine ehrlichiosis; canine tropical pancytopenia;
XX tick-borne rickettsial disease; serodiagnostics.

OS Ehrlichia canis.

XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= Signal_peptide
FT Protein 24..280
FT /label= Mature_Eca28SA3_28-kDa_protein

XX WO200032745-A2.

XX 08-JUN-2000.

XX 24-NOV-1999; 99WO-US28075.

XX 30-NOV-1998; 98US-0201458.

XX 03-MAR-1999; 99US-0261358.

XX (RERE-) RES DEV FOUND.

XX Walker DH, Yu X, McBride JW;

XX WPI: 2000-412298/35.

XX N-PSDB; AAD01294, AAD01295.
XX Ehrlichia canis antigens useful for vaccinating against canine
XX ehrlichiosis in dogs -

Claim 12; Page 68-69; 86pp; English.

XX The patent relates to homologous 28-kilobalton (kDa) protein genes of
XX Ehrlichia canis, designated Eca28SA1, Eca28SA2, Eca28SA3, Eca28-1 and
XX Eca28-2. These genes are members of a polymorphic multiple gene family
XX and contained in a single locus of 5.592 kb. The 28-kDa proteins are
XX immunoreactive with anti-E. canis serum hence are important
XX immunoprotective antigens. The protein is useful for vaccinating
XX against E. canis infections such as canine ehrlichiosis in dogs.
XX Canine ehrlichiosis, also known as canine tropical pancytopenia, is a
XX tick-borne rickettsial disease of dogs. Eca28-1 is conserved amongst
XX different strains of E. canis and hence useful for serodiagnostics of
XX canine ehrlichiosis. The present sequence is a E. canis
XX Eca28SA3 30-kDa protein which is post-translationally modified to a
XX mature 28-kDa protein by cleavage of N-terminal signal sequence.

XX Sequence 280 AA;

Query Match 71.8%; Score 79; DB 21; Length 280;
Best Local Similarity 82.4%; Pred. No. 7.8e-05;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
XXXXXXXXXXXXX

```
Db      60 STVGVEGLKHDWNGGTI 76
RESULT 10
AAU04198
ID      AAU04198 standard; Protein; 280 AA.
AC
XX
XX      AAU04198;
XX
XX      23-OCT-2001 (first entry)
XX
XX      Variable surface antigen 4 (VSA4) from Ehrlichia chaffeensis.
XX
XX      Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
XX      infection; heartwater; diagnostic; variable surface antigen; VSA.
XX
XX      Ehrlichia chaffeensis.
XX
XX      US6251872-B1.
XX
XX      26-JUN-2001.
XX
XX      17-OCT-1997; 97US-0953326.
XX
XX      17-OCT-1996; 96US-0733230.
XX
XX      (UYFL ) UNIV FLORIDA.
XX
XX      Barbet AF, Ganta RR, McGuire TC, Burrridge MJ, Nyika A;
XX      Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX
XX      WPI: 2001-424487/45.
XX      N-PSDB; AAS07578.
XX
XX      New MAP2 genes and polypeptides useful as vaccines for conferring
XX      immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX      as molecular markers in nucleic acid analysis procedures -
XX      Example 3; Fig 2A-2B; 30pp; English.
XX
XX      The sequence represents the amino acid sequence of variable surface
XX      antigen 4 (VSA4) isolated from Ehrlichia chaffeensis, which
XX      has similarity to major antigen protein (MAP). The MAP polynucleotides
XX      and polypeptides are useful as vaccines for conferring immunity to
XX      rickettsia infection, including Cowdria ruminantium causing heartwater.
XX      The MAP polynucleotides may be used as molecular markers in nucleic acid
XX      analysis procedures, and to produce the MAP polypeptides, which may
XX      be used to raise antibodies that are reactive with the polypeptides.
XX      The nucleic acids may further be used as probes to identify
XX      complementary sequences within other nucleic acid molecules or genomes,
XX      where such probes can be applied to identify or distinguish infectious
XX      strains of organisms in diagnostic procedures or in rickettsial
XX      research where identification of particular organisms or strains is
XX      needed.
XX
XX      Query Match 71.8%; Score 79; DB 22; Length 280;
XX      Best Local Similarity 73.7%; Pred. No. 7.8e-05;
XX      Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
XX
XX      Qy      2 STVGVEGLKHDWNGSPILK 20
XX      :||||| ||||| |
XX      Db      61 TTIGVEGLKQDWDGSTISK 79
XX
XX      RESULT 11
XX      ABG77953
XX      ID      ABG77953 standard; Protein; 280 AA.
XX      AC
XX      ABG77953;
XX
XX      15-NOV-2002 (first entry)
XX
XX      Ehrlichia canis outer membrane protein (P30F) #4.
XX      Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX
XX      Ehrlichia canis.
XX
XX      US2002120115-A1.
XX
XX      29-AUG-2002.
XX
XX      28-JAN-2002; 2002US-0059964.
XX
XX      19-MAY-1999; 99US-0314701.
XX      (RIKI/) RIKIHISA Y.
XX      (OHAS/) OHASHI N.
XX
XX      Rikihisa Y, Ohashi N;
XX
XX      WPI: 2002-618954/66.
XX      N-PSDB; ABS63294.
XX
XX      Isolated polynucleotide encoding an outer membrane protein of E. canis
XX      or E. chaffeensis used in the diagnosis of infection -
XX      Claim 10; Fig 2B; 49pp; English.
XX
XX      The invention relates to an isolated polynucleotide encoding an outer
XX      membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX      in the diagnosis of infection. An infection such as human ehrlichiosis or
XX      canine ehrlichiosis can be diagnosed by providing a serum sample from the
XX      patient, providing a polypeptide or mixture of polypeptides, contacting
XX      the sample with the polypeptide and assaying for the formation of a
XX      complex between antibodies in the serum sample and the polypeptide, where
XX      formation of a complex is indicative of infection with E. chaffeensis.
XX      This sequence represents an Ehrlichia outer membrane protein of the
XX      invention.
XX
XX      Query Match 71.8%; Score 79; DB 23; Length 280;
XX      Best Local Similarity 82.4%; Pred. No. 7.8e-05;
XX      Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX      Qy      2 STVGVEGLKHDWNGSPI 18
XX      ||||||||| |
XX      Db      60 STVGVEGLKHDWNGGTI 76
XX
XX      RESULT 12
XX      AAU96102
XX      ID      AAU96102 standard; Protein; 280 AA.
XX      AC
XX      AAU96102;
XX
XX      02-JUL-2002 (first entry)
XX
XX      Ehrlichia canis p28-6.
XX
XX      Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX      antibacterial.
XX
XX      Ehrlichia canis.
XX
XX      WO200222782-A2.
XX
XX      21-MAR-2002.
XX
XX      12-SEP-2001; 2001WO-US28759.
XX
XX      12-SEP-2000; 2000US-0660587.
XX
```

PA (RERE-) RES DEV FOUND.
XX Walker DH, Yu X, McBride JW;
XX
DR WPI: 2002-351882/38.
DR N-PSDB; ABK68854.
XX
XX New recombinant homologous 28 kilodalton immunodominant protein from
PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX
PS Claim 16; Figure 7; 106pp; English.
XX
CC The invention relates to a recombinant homologous 28 kDa immunodominant
CC protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
CC dispersed in a pharmaceutically acceptable carrier, is useful for
CC inhibiting E. canis infection in a subject. (I) is useful in the
CC development of vaccines and serodiagnostics that are particularly
CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
CC represent the 28-kDa antigen amino acid sequences of the invention.
XX
SQ Sequence 280 AA;

Query Match 71.8%; Score 79; DB 23; Length 280;
Best Local Similarity 82.4%; Pred. No. 7.8e-05;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSPILK 18
DB 60 STVGVGFLKHDWNGGTI 76
|||||

RESULT 13
AAU73417
ID AAU73417 standard; Protein; 280 AA.
XX
AC AAU73417;
XX
XX 12-MAR-2002 (first entry)
XX
DE Ehrlichia chaffeensis outer membrane protein P28-18.
XX
KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
XX
OS Ehrlichia chaffeensis.
XX
XX WO200183699-A2.
XX
XX 08-NOV-2001.
XX
XX 01-MAY-2001; 2001WO-US13997.
XX
XX 01-MAY-2000; 2000US-201035P.
XX
PA (RERE-) RES DEV FOUND.
XX Walker DH, Yu X;
XX
XX WPI: 2002-066527/09.
XX
XX Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
PT P28 useful as a vaccine against Ehrlichia chaffeensis -
XX
PS Disclosure; Figure 2; 97pp; English.
XX
XX The invention relates to isolated and purified 28-kDa outer membrane
CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
CC are encoded by a 28kDa outer membrane protein multigene family. P28
CC proteins are useful as a vaccine against E.chaffeensis, DNA encoding P28
CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
XX
XX Sequence 280 AA;

Query Match 71.8%; Score 79; DB 23; Length 280;
Best Local Similarity 73.7%; Pred. No. 7.8e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSPILK 20
DB 61 TTIGVGFLKQDWDGSTISK 79
:|||||

RESULT 14
AAY06948
ID AAY06948 standard; Protein; 280 AA.
XX
AC AAY06948;
XX
XX 05-JUL-1999 (first entry)
XX
XX E. chaffeensis OMP-1F protein.
DE
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
KW detection; dog.
XX
OS Ehrlichia chaffeensis.
XX
XX WO9913720-A1.
PN
XX 25-MAR-1999.
PD
XX 18-SEP-1998; 98WO-US19600.
PF
XX 19-SEP-1997; 97US-0059353.
PR
XX (OHIS) UNIV OHIO STATE.
PA
XX Ohashi N, Rikihisa Y;
PI
XX WPI: 1999-254290/21.
DR
XX N-PSDB; AAX34748.
DR
XX Novel outer membrane proteins from Ehrlichia chaffeensis and
PT Ehrlichia canis
XX
XX Claim 16; Fig 8B; 55pp; English.
PS
XX The invention provides isolated outer membrane proteins (OMP) from
CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX Sequence 280 AA;

Query Match 69.1%; Score 76; DB 20; Length 280;
Best Local Similarity 73.7%; Pred. No. 0.00025;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSPILK 20
DB 61 TTTGVGFLKQDWDGSTISK 79
:|||||

RESULT 15
ABG77940
ID ABG77940 standard; Protein; 280 AA.
XX
XX AC
XX ABG77940;
XX
XX 15-NOV-2002 (first entry)
DT
XX Ehrlichia chaffeensis outer membrane protein (OMP) #6.
DE
XX Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX

XX OS Ehrlichia chaffeensis.
 XX PN US2002120115-A1.
 XX PD 29-AUG-2002.
 XX PF 28-JAN-2002; 2002US-0059964.
 XX PR 19-MAY-1999; 99US-0314701.
 XX PA (RIKI/) RIKIHISA Y.
 XX PA (OHAS/) OHASHI N.
 XX PI Rikihisa Y, Ohashi N;
 XX WPI: 2002-618954/66.
 DR N-PSDB; ABS63281.
 XX Isolated polynucleotide encoding an outer membrane protein of E.canis
 PT or E.chaffeensis used in the diagnosis of infection -
 XX PS Disclosure; Fig 8B; 49pp; English.
 XX CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.
 XX Sequence 280 AA;
 SQ
 Query Match 69.1%; Score 76; DB 23; Length 280;
 Best Local Similarity 73.7%; Pred. No. 0.00025;
 Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 STVGVFGLKHDWDGSPILK 20
 : | | | | | | | | | |
 Db 61 TTGTVFGLKQDWDGSTISK 79
 RESULT 16
 ID AAU96110 standard; Protein; 280 AA.
 XX AAU96110;
 XX 02-JUL-2002 (first entry)
 XX Ehrlichia chaffeensis OMP-1F.
 XX Ehrlichia canis infection; vaccine; serodiagnostic; p28;
 XX antibacterial.
 XX Ehrlichia chaffeensis.
 XX W0200222782-A2.
 XX 21-MAR-2002.
 XX 12-SEP-2001; 2001WO-US28759.
 XX 12-SEP-2000; 2000US-0660587.
 XX (RERE-) RES DEV FOUND.
 XX Walker DH, Yu X, McBride JW;
 XX

DR WPI: 2002-351882/38.
 XX New recombinant homologous 28 kilodalton immunodominant protein from
 PT Ehrlichia canis, useful for treating Ehrlichia canis infections -
 XX Example 3; Figure 3; 106pp; English.
 XX The invention relates to a recombinant homologous 28 kDa immunodominant
 CC protein, p28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
 CC dispersed in a pharmaceutically acceptable carrier, is useful for
 CC inhibiting E. canis infection in a subject. (I) is useful in the
 CC development of vaccines and serodiagnostics that are particularly
 CC effective for disease prevention and serodiagnosis. AAU96100-AAU96118
 CC represent the 28-kDa antigen amino acid sequences of the invention.
 XX Sequence 280 AA;
 SQ
 Query Match 69.1%; Score 76; DB 23; Length 280;
 Best Local Similarity 73.7%; Pred. No. 0.00025;
 Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 STVGVFGLKHDWDGSPILK 20
 : | | | | | | | | | |
 Db 61 TTGTVFGLKQDWDGSTISK 79
 RESULT 17
 ID ABG30749 standard; Peptide; 19 AA.
 XX ABG30749;
 XX 21-OCR-2002 (first entry)
 XX Ehrlichia chaffeensis peptide fragment #5.
 XX Antibody detection; monoclonal antibody; polyclonal antibody.
 XX Ehrlichia chaffeensis.
 XX W0200257794-A2.
 XX 25-JUL-2002.
 XX 16-JAN-2002; 2002WO-US01395.
 XX 18-JAN-2001; 2001US-0765739.
 XX (IDEX-) IDEX LAB INC.
 XX Lawton R, O'Connor TP, Bartol BA, Machenry PS;
 XX WPI: 2002-599730/64.
 XX New composition of matter comprising a polypeptide, useful in detecting
 PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
 PT detecting or quantifying the presence of Ehrlichia infection in mammals
 PT
 PS Claim 1; Page 5; 29pp; English.
 XX The invention relates to a composition of matter comprising a polypeptide
 CC isolated from Ehrlichia species. The composition can be used for
 CC detecting the presence of antibodies to Ehrlichia, comprising contacting
 CC one or more polypeptides with a test sample suspected of comprising
 CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 CC complexes to form and detecting the complexes, where the detection of
 CC polypeptide/antibody complexes is an indication that antibodies to
 CC Ehrlichia are present in the test sample. The composition is useful for
 CC detecting or quantifying the presence of E. canis or E. chaffeensis
 CC infection in mammals. The polypeptides can be used to develop monoclonal
 CC and/or polyclonal antibodies that can be employed in assay systems and in
 CC the generation of chimeric antibodies for therapeutic use or other

CC similar applications. This sequence represents an E. chaffeensis peptide
XX fragment used in the composition of the invention.

SQ Sequence 19 AA;

Query Match 66.4%; Score 73; DB 23; Length 19;
Best Local Similarity 76.5%; Pred. No. 3.4e-05;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
Db 2 TTGVGFLKQNDWDGST1 18

RESULT 18

ABG30745

ID ABG30745 standard; Peptide; 20 AA.

XX AC ABG30745;

XX DT 21-OCT-2002 (first entry)

XX DE Ehrlichia chaffeensis peptide fragment #1.

XX KW Antibody detection; monoclonal antibody; polyclonal antibody.

XX OS Ehrlichia chaffeensis.

XX PN WO200257794-A2.

XX PD 25-JUL-2002.

XX PF 16-JAN-2002; 2002WO-US01395.

XX PR 18-JAN-2001; 2001US-0765739.

XX PA (IDEX-) IDEXX LAB INC.

XX PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;

XX DR WPI; 2002-599730/64.

XX PT New composition of matter comprising a polypeptide, useful in detecting
PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
PT detecting or quantifying the presence of Ehrlichia infection in mammals

PS Claim 1; Page 5; 29pp; English.

CC The invention relates to a composition of matter comprising a polypeptide
CC isolated from Ehrlichia species. The composition can be used for
CC detecting the presence of antibodies to Ehrlichia, comprising contacting
CC one or more polypeptides with a test sample suspected of comprising
CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
CC complexes to form and detecting the complexes, where the detection of
CC polypeptide/antibody complexes is an indication that antibodies to
CC Ehrlichia are present in the test sample. The composition is useful for
CC detecting or quantifying the presence of E. canis or E. chaffeensis
CC infection in mammals. The polypeptides can be used to develop monoclonal
CC and/or polyclonal antibodies that can be employed in assay systems and in
CC the generation of chimeric antibodies for therapeutic use or other
CC similar applications. This sequence represents an E. chaffeensis peptide
CC fragment used in the composition of the invention.

SQ Sequence 20 AA;

Query Match 65.5%; Score 72; DB 23; Length 20;
Best Local Similarity 76.5%; Pred. No. 5.3e-05;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
Db 2 TTGVGFLKQNDWDGSAI 18

RESULT 19

AAY06942

ID AAY06942 standard; Protein; 256 AA.

XX AC AAY06942;

XX DT 05-JUL-1999 (first entry)

XX DE E. chaffeensis p28 protein.

XX KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.

XX OS Ehrlichia chaffeensis.

XX PN WO9913720-A1.

XX PD 25-MAR-1999.

XX PF 18-SEP-1998; 98WO-US19600.

XX PR 19-SEP-1997; 97US-0059353.

XX PA (OHIS) UNIV OHIO STATE.

XX PI Ohashi N, Rikihisa Y;

XX DR WPI; 1999-254290/21.

XX DR N-PSDB; AAX34742.

XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and

XX PT Ehrlichia canis

XX PS Claim 18; Fig 1; 55pp; English.

XX CC The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the P30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.

SQ Sequence 256 AA;

Query Match 65.5%; Score 72; DB 20; Length 256;
Best Local Similarity 76.5%; Pred. No. 0.001;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18

Db 35 TTGVGFLKQNDWDGSAI 51

RESULT 20

ABG77966

ID ABG77966 standard; Protein; 256 AA.

XX AC ABG77966;

XX DT 15-NOV-2002 (first entry)

XX DE Protein encoded by Ehrlichia chaffeensis p28 gene.

XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection; p28; OMP-1.

XX OS Ehrlichia chaffeensis.

XX PN US2002120115-A1.

XX PD 29-AUG-2002.

PF 28-JAN-2002; 2002US-0059964.
 PR 19-MAY-1999; 99US-0314701.
 PA (RIKI/) RIKIHISA Y.
 PA (OHASHI/) OHASHI N.
 XX Rikihisa Y, Ohashi N;
 XX WPI: 2002-618954/66.
 DR N-PSDB; ABS63307.
 XX
 XX Isolated polynucleotide encoding an outer membrane protein of E.coli
 PT or E.chaffeensis used in the diagnosis of infection -
 PS Disclosure; Fig 1; 49pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents the Ehrlichia chaffeensis OMP-1 protein encoded
 CC by the p28 gene.
 XX
 XX Sequence 256 AA;
 SQ
 Query Match 65.5%; Score 72; DB 23; Length 256;
 Best Local Similarity 76.5%; Pred. No. 0.001;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 STGVGFLKHDWDGSP1 18
 Db :||||| :|||||
 35 TTGVGFLKQNDGSAI 51
 RESULT 21
 AAW51095
 ID AAW51095 standard; Protein; 276 AA.
 XX AC AAW51095;
 XX
 XX 14-SEP-1998 (first entry)
 DE Ehrlichia chaffeensis VSA5 protein (partial sequence).
 XX MAP1 homologue; variable surface antigen; VSA5; rickettsia;
 KW DNA vaccine.
 XX Ehrlichia chaffeensis.
 OS
 XX Key Location/Qualifiers
 FH Peptide 1..25
 FT /note= "putative signal peptide"
 FT
 FW W09816554-A1.
 XX
 PD 23-APR-1998.
 XX
 XX 17-OCT-1997; 97WO-US19044.
 XX
 XX 17-OCT-1996; 96US-0733230.
 XX
 XX (UYFL) UNIV FLORIDA.
 PA
 XX Barbet AF, Burrig MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX WPI: 1998-251232/22.
 DR N-PSDB; AAV07179.
 XX

XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 XX Claim 3; Fig 2B; 39pp; English.
 XX
 CC This is the near full-length variable surface antigen VSA5 protein
 CC of Ehrlichia chaffeensis; it lacks about 5-7 C-terminal amino acid
 CC residues. The VSA5 amino acid sequence was deduced from a partial
 CC open reading frame of a genomic locus (see AAV07179) of E.
 CC chaffeensis that was obtained on the basis of homology to the major
 CC antigenic protein MAP1 (see AAW51088) of Cowdria ruminantium. This
 CC genomic locus included 5 ORFs encoding similar, but non-identical
 CC proteins (see AAW51091-95). A claimed composition comprises a
 CC nucleic acid (see AAV07176-82) encoding a polypeptide (see AAW51088-99)
 CC that elicits a protective immune response against a rickettsial
 CC pathogen. The nucleic acid is used, in human or veterinary
 CC medicine, in vaccines to protect against Rickettsia, Ehrlichia,
 CC Anaplasma and Cowdria species. The Ehrlichia antigenic
 CC polypeptides can be used diagnostically to detect antibodies
 CC associated with Ehrlichia infection (claimed).
 XX
 XX Sequence 276 AA;
 SQ
 Query Match 65.5%; Score 72; DB 19; Length 276;
 Best Local Similarity 76.5%; Pred. No. 0.0011;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 STGVGFLKHDWDGSP1 18
 Db :||||| :|||||
 60 TTGVGFLKQNDGSAI 76
 RESULT 22
 AAB36189
 ID AAB36189 standard; Protein; 276 AA.
 XX AC AAB36189;
 XX
 XX 02-MAR-2001 (first entry)
 DE Ehrlichia chaffeensis partial VSA5.
 XX Ehrlichia chaffeensis; VSA5; variable surface antigen 5; MAP1;
 KW major antigenic protein 1; antirickettsial; vaccine; gene therapy;
 KW Rickettsia; Cowdria; Anaplasma; map2; ihworf3; 4hworf1; 18hworf1;
 XX 39dorf3.
 OS Ehrlichia chaffeensis.
 XX
 XX W0200065063-A2.
 PN
 PD 02-NOV-2000.
 XX
 XX 21-APR-2000; 2000WO-US10886.
 XX
 XX 22-APR-1999; 99US-0130725.
 XX
 XX (UYFL) UNIV FLORIDA.
 PA
 XX Barbet AF, Bowie MV, Ganta RR, Burrig MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmore WW, Allemen AR;
 XX WPI: 2000-679675/66.
 DR N-PSDB; AAC68706.
 XX
 XX New polynucleotides useful as DNA vaccines for conferring immunity to
 PT rickettsial infection e.g. heartwater caused by Cowdria ruminantium
 PT comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
 XX
 XX Claim 3; Page 47; 63pp; English.
 PS

XX The present sequence shows a high degree of similarity to the major
 CC antigenic protein 1 (MAP1) of Ehrlichia sp. The MAP1 gene may be
 CC used in a vaccine to protect animals or humans against rickettsial
 CC diseases caused by a organisms of Rickettsia sp., Ehrlichia sp.,
 CC Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 CC protective against the rickettsial pathogen. The nucleic acid vaccines
 CC can be driven by the human cytomegalovirus (HCMV) enhancer-promoter.
 CC Cowdria ruminatum genes designated map 2, lhworf3, 4hworf1, 18hworf1
 CC and 3gdrf3 may be used in therapeutic and diagnostic applications. The
 CC polypeptides are useful for detecting antibodies associated with
 CC infection by a rickettsial pathogen whilst the polynucleotides may be
 CC used to detect the presence of rickettsial nucleic acids.

XX Query Match 65.5%; Score 72; DB 21; Length 276;
 CC Best Local Similarity 76.5%; Pred. No. 0.0011;
 CC Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
 DB :|||||||:|||||
 60 TTVGVFLKQNWDSAI 76

RESULT 23
 AAU04199
 ID AAU04199 standard; Protein; 276 AA.
 AC AAU04199;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Variable surface antigen 5 (VSA5) from Ehrlichia chaffeensis.
 KW Major antigenic protein; MAP; vaccine; immunogenic; rickettsia;
 KW infection; heartwater; diagnostic; variable surface antigen; VSA.
 XX Ehrlichia chaffeensis.
 OS US6251872-B1.
 PN Rurangirwa FR, Mahan SM, Bowlie MV, Alleman AR;
 XX
 PD 26-JUN-2001.
 XX
 PF 17-OCT-1997; 97US-0953326.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Ganta RR, McGuire TC, Burridge MJ, Nyika A;
 PI Rurangirwa FR, Mahan SM, McGuire TC, Burridge MJ, Nyika A;
 XX
 DR WPI: 2001-424487/45.
 DR N-PSDB; AAS07578.
 XX
 PT New MAP2 genes and polypeptides useful as vaccines for conferring
 PT immunity to human and animal rickettsial diseases, e.g. heartwater, or
 PT as molecular markers in nucleic acid analysis procedures -
 XX
 PS Example 3; Fig 2A-2B; 30pp; English.
 XX
 CC The sequence represents the amino acid sequence of variable surface
 CC antigen 5 (VSA5) isolated from Ehrlichia chaffeensis, which
 CC has similarity to major antigen protein (MAP). The MAP polynucleotides
 CC and polypeptides are useful as vaccines for conferring immunity to
 CC rickettsia infection, including Cowdria ruminatum causing heartwater.
 CC The MAP polynucleotides may be used as molecular markers in nucleic acid
 CC analysis procedures, and to produce the MAP polypeptides, which may
 CC be used to raise antibodies that are reactive with the polypeptides.
 CC The nucleic acids may further be used as probes to identify
 CC complementary sequences within other nucleic acid molecules or genomes,
 CC where such probes can be applied to identify or distinguish infectious

CC strains of organisms in diagnostic procedures or in rickettsial
 CC research where identification of particular organisms or strains is
 CC needed.

XX Query Match 65.5%; Score 72; DB 22; Length 276;
 CC Best Local Similarity 76.5%; Pred. No. 0.0011;
 CC Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
 DB :|||||||:|||||
 60 TTVGVFLKQNWDSAI 76

RESULT 24
 AAU51089
 ID AAU51089 standard; Protein; 280 AA.
 XX
 AC AAU51089;
 XX
 DT 14-SEP-1998 (first entry)
 XX
 DE Ehrlichia chaffeensis major antigenic protein 1 (MAP1).
 XX MAP1 gene; major antigenic protein 1; rickettsia; DNA vaccine.
 KW Ehrlichia chaffeensis.
 OS WO9816554-A1.
 PN
 XX
 PD 23-APR-1998.
 XX
 PF 17-OCT-1997; 97WO-US19044.
 XX
 PR 17-OCT-1996; 96US-0733230.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Burridge MJ, Ganta RR, Mahan SM, McGuire TC;
 PI Nyika A, Rurangirwa FR;
 XX
 DR WPI: 1998-251232/22.
 DR N-PSDB; AAV07177.
 XX
 PT Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals
 XX
 PS Claim 3; Page 18-19; 39pp; English.
 XX
 CC This polypeptide comprises the major antigen protein 1 gene (MAP1)
 CC of Ehrlichia chaffeensis. It is encoded by the MAP1 gene (see
 CC AAV07177). A claimed composition comprises a nucleic acid (see
 CC AAV07176-82) encoding a polypeptide (see AAU51088-99) that elicits a
 CC protective immune response against a rickettsial pathogen. The
 CC nucleic acid is used, in human or veterinary medicine, in vaccines
 CC to protect against Rickettsia, Ehrlichia, Anaplasma and Cowdria
 CC species. The nucleic acid does not replicate in the host but
 CC remains episomal and capable of expressing polypeptide for at least
 CC 19 mth. The Ehrlichia antigenic polypeptides can be used
 CC diagnostically to detect antibodies associated with Ehrlichia
 CC infection (claimed).
 XX
 SQ Sequence 280 AA;

Query Match 65.5%; Score 72; DB 19; Length 280;
 CC Best Local Similarity 76.5%; Pred. No. 0.0011;
 CC Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVGFLKHDWDGSP1 18
 DB :|||||||:|||||
 61 TTVGVFLKQNWDSAI 77

```

RESULT 25
AAB36183
ID AAB36183 standard; Protein; 280 AA.
AC AAB36183;
DT 02-MAR-2001 (first entry)
DE Ehrlichia chaffeensis MAP1.
KW Ehrlichia chaffeensis; MAP1; major antigenic protein 1; antirickettsial;
KW vaccine; gene therapy; Rickettsia; Cowdria; Anaplasma; map2; lhworf3;
KW 4hworf1; lhworf1; 5gdorf3.
OS Ehrlichia chaffeensis.
XX WO200065063-A2.
XX 02-NOV-2000.
XX 21-APR-2000; 2000WO-US10886.
XX 22-APR-1999; 99US-0130725.
XX (UYFL) UNIV FLORIDA.
XX Barbet AF, Bowie MV, Ganta RR, Burrigide MJ, Mahan SM, McGuire TC;
XX Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
XX WPI; 2000-679675/66.
XX N-PSDB; AAC68700.
XX New polynucleotides useful as DNA vaccines for conferring immunity to
XX rickettsial infection e.g. heartwater caused by Cowdria ruminantium.
XX comprises major antigenic protein 1 or 2 gene of rickettsial pathogens
XX Claim 3; Page 35-36; 63pp; English.
XX The present sequence is given in a specification relating to nucleic
XX acid vaccines containing genes to protect animals or humans against
XX rickettsial diseases caused by a organism of Rickettsia sp., Ehrlichia
XX sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
XX protective against the rickettsial pathogen. The vaccine comprises the
XX major antigenic protein 1 (MAP1) gene of the major antigenic protein 2
XX (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
XX driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
XX ruminantium genes designated map 2, lhworf3, 4hworf1, lhworf1 and
XX 5gdorf3 may be used in therapeutic and diagnostic applications. The
XX polypeptides are useful for detecting antibodies associated with
XX infection by a rickettsial pathogen whilst the polynucleotides may be
XX used to detect the presence of rickettsial nucleic acids.
XX Sequence 280 AA;
XX Query Match 65.5%; Score 72; DB 21; Length 280;
XX Best Local Similarity 76.5%; Pred. No. 0.0011;
XX Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 STVGVLKHDWDGSP1 18
DB 61 TTGVVFLKQNDGSAI 77
RESULT 26
AAU04193
ID AAU04193 standard; Protein; 280 AA.
AC AAU04193;
XX 23-OCT-2001 (first entry)

```

```

XX Major antigenic protein 1 (MAP1) from Ehrlichia chaffeensis.
DE Major antigenic protein 1; MAP1; vaccine; immunogenic; rickettsia;
XX infection; heartwater; diagnostic.
KW Ehrlichia chaffeensis.
XX US6251872-B1.
XX 26-JUN-2001.
XX 17-OCT-1997; 97US-0953326.
XX 17-OCT-1996; 96US-0733230.
XX (UYFL) UNIV FLORIDA.
XX Barbet AF, Ganta RR, McGuire TC, Burrigide MJ, Nyika A;
XX Rurangirwa FR, Mahan SM, Bowie MV, Alleman AR;
XX WPI; 2001-424487/45.
XX N-PSDB; AAS07576.
XX New MAP2 genes and polypeptides useful as vaccines for conferring
XX immunity to human and animal rickettsial diseases, e.g. heartwater, or
XX as molecular markers in nucleic acid analysis procedures.
XX Disclosure; Column 15-17; 30pp; English.
XX The sequence represents the amino acid sequence of major antigenic
XX protein 1 (MAP1) from Ehrlichia chaffeensis. The MAP polynucleotides and
XX polypeptides are useful as vaccines for conferring immunity to rickettsia
XX infection, including Cowdria ruminantium causing heartwater. The MAP
XX polynucleotides may be used as molecular markers in nucleic acid
XX analysis procedures, and to produce the MAP polypeptides, which may
XX be used to raise antibodies that are reactive with the polypeptides.
XX The nucleic acids may further be used as probes to identify
XX complementary sequences within other nucleic acid molecules or genomes,
XX where such probes can be applied to identify or distinguish infectious
XX strains of organisms in diagnostic procedures or in rickettsial
XX research where identification of particular organisms or strains is
XX needed.
XX Sequence 280 AA;
XX Query Match 65.5%; Score 72; DB 22; Length 280;
XX Best Local Similarity 76.5%; Pred. No. 0.0011;
XX Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 STVGVLKHDWDGSP1 18
DB 61 TTGVVFLKQNDGSAI 77
RESULT 27
AAU06943
ID AAU06943 standard; Protein; 281 AA.
AC AAU06943;
XX 05-JUL-1999 (first entry)
XX E. chaffeensis OMP-1 protein.
XX Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
XX detection; dog.
XX Ehrlichia chaffeensis.
XX WO9913720-A1.
XX 25-MAR-1999.

```

XX PF 18-SEP-1998; 98WO-US19600.
XX PR 19-SEP-1997; 97US-0059353.
XX PA (OHIS) UNIV OHIO STATE.
XX PI Ohashi N, Rikihisa Y;
XX WPI: 1999-254290/21.
XX DR N-PSDB; AAX34743.
XX PT Novel outer membrane proteins from Ehrlichia chaffeensis and
XX Ehrlichia canis
XX PS Disclosure; Fig 3B; 55pp; English.
XX CC The invention provides isolated outer membrane proteins (OMP) from
XX Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
XX of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
XX in AAY06943-958. The E. canis proteins form part of the p30 family and
XX consist of proteins shown in AAY06959-970. The proteins and genes are
XX used to detect E. chaffeensis in patients and E. canis in dogs.
XX SQ Sequence 281 AA;
Query Match 65.5%; Score 72; DB 20; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.0011;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 STVGVGFLKHDWDGSP1 18
Db 60 TTVGVFGLKQNWGSAI 76
RESULT 28
ABG77935
ID ABG77935 standard; Protein; 281 AA.
XX AC ABG77935;
XX DT 15-NOV-2002 (first entry)
XX DE Ehrlichia chaffeensis outer membrane protein (OMP) #1.
XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.
XX OS Ehrlichia chaffeensis.
XX US2002120115-A1.
XX PD 29-AUG-2002.
XX PF 28-JAN-2002; 2002US-0059964.
XX PR 19-MAY-1999; 99US-0314701.
XX PA (RIKI/) RIKIHISA Y.
XX PA (OHAS/) OHASHI N.
XX PI Rikihisa Y, Ohashi N;
XX WPI: 2002-618954/66.
XX DR N-PSDB; ABS63276.
XX PT Isolated polynucleotide encoding an outer membrane protein of E.canis
XX or E.chaffeensis used in the diagnosis of infection -
XX PS Claim 14; Fig 3B; 49pp; English.
XX CC The invention relates to an isolated polynucleotide encoding an outer
XX membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
XX in the diagnosis of infection. An infection such as human ehrlichiosis or

CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
CC patient, providing a polypeptide or mixture of polypeptides, contacting
CC the sample with the polypeptide and assaying for the formation of a
CC complex between antibodies in the serum sample and the polypeptide, where
CC formation of a complex is indicative of infection with E. chaffeensis.
CC This sequence represents an Ehrlichia outer membrane protein of the
CC invention.
XX SQ Sequence 281 AA;
Query Match 65.5%; Score 72; DB 23; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.0011;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 STVGVGFLKHDWDGSP1 18
Db 60 TTVGVFGLKQNWGSAI 76
RESULT 29
AAU96105
ID AAU96105 standard; Protein; 281 AA.
XX AC AAU96105;
XX DT 02-JUL-2002 (first entry)
XX DE Ehrlichia chaffeensis P28.
XX KW Ehrlichia canis infection; vaccine; serodiagnostic; p28;
XX KW antibacterial.
XX OS Ehrlichia chaffeensis.
XX PN WO200222782-A2.
XX PD 21-MAR-2002.
XX PF 12-SEP-2001; 2001WO-US28759.
XX PR 12-SEP-2000; 2000US-0660587.
XX PA (RERE-) RES DEV FOUND.
XX PI Walker DH, Yu X, McBride JW;
XX WPI: 2002-351882/38.
XX PT New recombinant homologous 28 kilodalton immunodominant protein from
XX Ehrlichia canis, useful for treating Ehrlichia canis infections -
XX Example 3; Figure 3; 106pp; English.
XX CC The invention relates to a recombinant homologous 28 kDa immunodominant
XX protein, P28, (I), of Ehrlichia canis. (I), a 28-kDa antigen preferably
XX dispersed in a pharmaceutically acceptable carrier, is useful for
XX inhibiting E. canis infection in a subject. (I) is useful in the
XX development of vaccines and serodiagnostics that are particularly
XX effective for disease prevention and serodiagnosis. AAU96100-AAU96118
XX represent the 28-kDa antigen amino acid sequences of the invention.
XX SQ Sequence 281 AA;
Query Match 65.5%; Score 72; DB 23; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.0011;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 STVGVGFLKHDWDGSP1 18
Db 60 TTVGVFGLKQNWGSAI 76
RESULT 30

AAU73418
 TD AAU73418 standard; Protein; 281 AA.
 XX
 AC AAU73418;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Ehrlichia chaffeensis outer membrane protein P28-19.
 XX
 KW Ehrlichia; outer membrane protein; P28; antibiotic; vaccine.
 XX
 OS Ehrlichia chaffeensis.
 XX
 PN WO200183699-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 01-MAY-2001; 2001WO-US13997.
 XX
 PR 01-MAY-2000; 2000US-201035P.
 XX
 PA (RERE-) RES DEV FOUND.
 XX
 PI Walker DH, Yu X;
 XX
 DR WPI; 2002-066527/09.
 XX
 PT Novel Ehrlichia chaffeensis 28-kDa outer membrane protein, designated
 PT P28 useful as a vaccine against Ehrlichia chaffeensis
 XX
 PS Disclosure; Figure 2; 97pp; English.
 XX
 CC The invention relates to isolated and purified 28-kDa outer membrane
 CC proteins (P28-1 to P28-21) of Ehrlichia chaffeensis. P28 proteins
 CC are encoded by a 28kDa outer membrane protein multigene family. P28
 CC proteins are useful as a vaccine against E. chaffeensis. DNA encoding P28
 CC is useful for transfecting a host cell. AAU73400-AAU73420 represent
 CC Ehrlichia chaffeensis P28 outer membrane proteins of the invention.
 XX
 SQ Sequence 281 AA;
 Query Match 65.5%; Score 72; DB 23; Length 281;
 Best Local Similarity 76.5%; Pred. No. 0.0011;
 Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 2 STGVGFLKQWDGSAI 18
 Db :|||||:|||||
 60 TTGVGFLKQWDGSAI 76
 RESULT 31
 ABG30744
 ID ABG30744 standard; Peptide; 20 AA.
 XX
 AC ABG30744;
 XX
 DT 21-OCT-2002 (first entry)
 XX
 DE Ehrlichia canis peptide fragment #2.
 XX
 KW Antibody detection; monoclonal antibody; polyclonal antibody.
 XX
 OS Ehrlichia canis.
 XX
 PN WO200257794-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 16-JAN-2002; 2002WO-US01395.
 XX
 PR 18-JAN-2001; 2001US-0765739.
 XX
 PA (IDEX-) IDEX LAB INC.

XX
 PI Lawton R, O'Connor TP, Bartol BA, Machenry PS;
 XX WPI; 2002-599730/64.
 DR
 XX
 PT New composition of matter comprising a polypeptide, useful in detecting
 PT the presence of antibodies to Ehrlichia canis or chaffeensis, or in
 PT detecting or quantifying the presence of Ehrlichia infection in mammals
 XX
 PS Claim 1; Page 5; 29pp; English.
 XX
 CC The invention relates to a composition of matter comprising a polypeptide
 CC isolated from Ehrlichia species. The composition can be used for
 CC detecting the presence of antibodies to Ehrlichia, comprising contacting
 CC one or more polypeptides with a test sample suspected of comprising
 CC antibodies to Ehrlichia, under conditions that allow polypeptide/antibody
 CC complexes to form and detecting the complexes, where the detection of
 CC polypeptide/antibody complexes is an indication that antibodies to
 CC Ehrlichia are present in the test sample. The composition is useful for
 CC infection in quantifying the presence of E. canis or E. chaffeensis
 CC and/or polyclonal antibodies that can be employed in assay systems and in
 CC the generation of chimeric antibodies for therapeutic use or other
 CC similar applications. This sequence represents an E. canis peptide
 CC fragment used in the composition of the invention.
 XX
 SQ Sequence 20 AA;
 Query Match 63.6%; Score 70; DB 23; Length 20;
 Best Local Similarity 70.6%; Pred. No. 0.00011;
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 2 STGVGFLKQWDGSP 18
 Db :|||||:|||||
 2 TTGVGFLKQWDGATI 18
 RESULT 32
 AAY06959
 ID AAY06959 standard; Protein; 288 AA.
 XX
 AC AAY06959;
 XX
 DT 05-JUL-1999 (first entry)
 XX
 DE E. canis P30 protein.
 XX
 KW Outer membrane protein; OMP; Ehrlichia chaffeensis; E. canis; P30;
 KW detection; dog.
 XX
 OS Ehrlichia canis.
 XX
 PN WO9913720-A1.
 XX
 PD 25-MAR-1999.
 XX
 PF 18-SEP-1998; 98WO-US19600.
 XX
 PR 19-SEP-1997; 97US-0059353.
 XX
 PA (OHIS) UNIV OHIO STATE.
 XX
 PI Ohashi N, Rikihisa Y;
 XX
 DR WPI; 1999-254290/21.
 XX
 DR N-PSDB; AAX34759.
 XX
 PT Novel outer membrane proteins from Ehrlichia chaffeensis and
 PT Ehrlichia canis
 XX
 PS Disclosure; Fig 19B; 55pp; English.

CC The invention provides isolated outer membrane proteins (OMP) from
 CC Ehrlichia chaffeensis and E. canis. The E. chaffeensis proteins form part
 CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
 CC in AAV06943-958. The E. canis proteins form part of the P30 family and
 CC consist of proteins shown in AAV06959-970. The proteins and genes are
 CC used to detect E. chaffeensis in patients and E. canis in dogs.

XX Sequence 288 AA;

Query Match 63.6%; Score 70; DB 20; Length 288;
 Best Local Similarity 70.6%; Pred. No. 0.0025; 3; Indels 0; Gaps 0;
 Matches 12; Conservative 2; Mismatches 3;

QY 2 STVGVEGLKHDWDGSP1 18
 : ||||| ||||| :
 Db 61 TTGTGVEGLKQDWDGATI 77

RESULT 33

ABG77950
 ID ABG77950 standard; Protein; 288 AA.

XX AC ABG77950;

XX DT 15-NOV-2002 (first entry)

XX DE Ehrlichia canis outer membrane protein (P30F) #1.

XX KW Outer membrane protein; OMP; P30F; ehrlichiosis; infection.

XX OS Ehrlichia canis.

XX PN US2002120115-A1.

XX PD 29-AUG-2002.

XX PF 28-JAN-2002; 2002US-0059964.

XX PR 19-MAY-1999; 99US-0314701.

XX PA (RIKL/) RIKIHISA Y.

XX PA (OHAS/) OHASHI N.

XX PI Rikihisa Y, Ohashi N;

XX DR WPI; 2002-618954/66.

XX DR N-PSDB; ABS63291.

XX Isolated polynucleotide encoding an outer membrane protein of E.canis
 PT or E.chaffeensis used in the diagnosis of infection -

XX Claim 10; Fig 19B; 49pp; English.

XX The invention relates to an isolated polynucleotide encoding an outer
 CC membrane protein (OMP) of Ehrlichia canis or Ehrlichia chaffeensis used
 CC in the diagnosis of infection. An infection such as human ehrlichiosis or
 CC canine ehrlichiosis can be diagnosed by providing a serum sample from the
 CC patient, providing a polypeptide or mixture of polypeptides, contacting
 CC the sample with the polypeptide and assaying for the formation of a
 CC complex between antibodies in the serum sample and the polypeptide, where
 CC formation of a complex is indicative of infection with E. chaffeensis.
 CC This sequence represents an Ehrlichia outer membrane protein of the
 CC invention.

XX Sequence 288 AA;

Query Match 63.6%; Score 70; DB 23; Length 288;
 Best Local Similarity 70.6%; Pred. No. 0.0025;
 Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
 : ||||| ||||| :
 Db 61 TTGTGVEGLKQDWDGATI 77

RESULT 34

AAW51088

ID AAW51088 standard; Protein; 287 AA.

XX AC AAW51088;

XX DT 14-SEP-1998 (first entry)

XX DE Cowdria ruminantium major antigenic protein 1 (MAP1).

XX KW MAP1 gene; major antigenic protein 1; rickettsia; heartworm;
 XX DNA vaccine.

XX OS Cowdria ruminantium.

XX PN W09816554-A1.

XX PD 23-APR-1998.

XX PF 17-OCT-1997; 97WO-US19044.

XX PR 17-OCT-1996; 96US-0733230.

XX PA (UYFL) UNIV FLORIDA.

XX PI Barbet AF, Burr ridge MJ, Ganta RR, Mahan SM, McGuire TC;

XX PI Nyika A, Rurangirwa FR;

XX DR WPI; 1998-251232/22.

XX DR N-PSDB; AAV07176.

XX Composition containing nucleic acid encoding rickettsial antigen -
 PT useful for, e.g. stimulating protective immune response in humans or
 PT animals

XX Claim 3; Page 15-16; 39pp; English.

XX This polypeptide comprises the major antigen protein 1 gene (MAP1)
 CC of Cowdria ruminantium, the causative agent of heartwater in
 CC domestic ruminants. It is encoded by the MAP1 gene (see AAV07176).
 CC A claimed composition comprises a nucleic acid (see AAV07176-82)
 CC encoding a polypeptide (see AAW51088-99) that elicits a protective
 CC immune response against a rickettsial pathogen. The nucleic acid
 CC is used, in human or veterinary medicine, in vaccines to protect
 CC against Rickettsia, Ehrlichia, Anaplasma and Cowdria species. The
 CC nucleic acid does not replicate in the host but remains episomal
 CC and capable of expressing polypeptide for at least 19 mth.

XX Sequence 287 AA;

Query Match 54.5%; Score 60; DB 19; Length 287;
 Best Local Similarity 73.3%; Pred. No. 0.12;
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGVEGLKHDWDG 15
 |:| ||||| |||||
 Db 61 KNTQTVEGLKQDWDG 75

RESULT 35

AAB36182

ID AAB36182 standard; Protein; 287 AA.

XX AC AAB36182;

XX DT 02-MAR-2001 (first entry)

XX DE Cowdria ruminantium MAP1.

XX Cowdria ruminantium; MAP1; major antigenic protein 1; antirickettsial;
 KW vaccine; gene therapy; Rickettsia; Ehrlichia; Anaplasma; map2; lhworf3;

KW 4hworfl; 18hworfl; 3gdorf3.
 XX
 OS Cowdria ruminantium.
 XX
 PN WO2000065063-A2.
 XX
 PD 02-NOV-2000.
 XX
 PF 21-APR-2000; 2000WO-US10886.
 XX
 PR 22-APR-1999; 99US-0130725.
 XX
 PA (UYFL) UNIV FLORIDA.
 XX
 PI Barbet AF, Bowie MV, Ganta RR, Burridge MJ, Mahan SM, McGuire TC;
 PI Rurangirwa FR, Moreland AL, Simbi BH, Whitmire WW, Alleman AR;
 DR WPI; 2000-679675/66.
 DR N-PSDB; AAC68699.
 XX

New polynucleotides useful as DNA vaccines for conferring immunity to
 rickettsial infection e.g. heartwater caused by cowdria ruminantium,
 comprises major antigenic protein 1 or 2 gene of rickettsial pathogens

Claim 3; Page 33; 63pp; English.

The present sequence is given in a specification relating to nucleic
 acid vaccines containing genes to protect animals or humans against
 rickettsial diseases caused by a organism of Rickettsia sp., Ehrlichia
 sp., Anaplasma sp. or Cowdria sp. The vaccine elicits an immune response
 protective against the rickettsial pathogen. The vaccine comprises the
 major antigenic protein 1 (MAP1) gene or the major antigenic protein 2
 (MAP2) gene of rickettsial pathogens. The nucleic acid vaccines can be
 driven by the human cytomegalovirus (HCMV) enhancer-promoter. Cowdria
 ruminantium genes designated map 2, 1hworfl, 18hworfl and
 3gdorf3 may be used in therapeutic and diagnostic applications. The
 polypeptides are useful for detecting antibodies associated with
 infection by a rickettsial pathogen whilst the polynucleotides may be
 used to detect the presence of rickettsial nucleic acids.

SO Sequence 287 AA;
 Query Match 54.5%; Score 60; DB 21; Length 287;
 Best Local Similarity 73.3%; Pred. No. 0.12;
 Matches 1; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 KSTGCVFGLKHDWDG 15
 Db :||| ||||| |||||
 61 KNTQTVFGLKRDWDG 75

Search completed: October 6, 2003, 07:47:55
 Job time : 38.3731 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 ; search time 11.0597 Seconds
(without alignments)
72.688 Million cell updates/sec

Title: US-09-765-739A-7

Perfect score: 103

Sequence: 1 NTUUGVFLKQDWDGSTIS 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103	100.0	280	4	US-09-660-587-14
2	103	100.0	280	4	US-09-261-358A-14
3	103	100.0	280	4	US-09-201-458-10
4	103	100.0	280	4	US-09-314-701-12
5	97	94.2	280	3	US-08-953-326-17
6	96	93.2	288	4	US-09-314-701-32
7	88	85.4	276	3	US-08-953-326-18
8	88	85.4	280	3	US-08-733-230-4
9	88	85.4	280	3	US-08-953-326-4
10	88	85.4	281	4	US-09-660-587-9
11	88	85.4	281	4	US-09-261-358A-9
12	88	85.4	281	4	US-09-201-458-5
13	88	85.4	281	4	US-09-314-701-2
14	80	77.7	280	4	US-09-660-587-6
15	80	77.7	280	4	US-09-261-358A-6
16	80	77.7	280	4	US-09-314-701-38
17	74	71.8	286	3	US-08-953-326-15
18	74	71.8	286	4	US-09-660-587-12
19	74	71.8	286	4	US-09-261-358A-12
20	74	71.8	286	4	US-09-201-458-8
21	74	71.8	286	4	US-09-314-701-8
22	73	70.9	278	4	US-09-660-587-2
23	73	70.9	278	4	US-09-261-358A-2
24	73	70.9	278	4	US-09-201-458-2
25	73	70.9	307	4	US-09-314-701-36
26	72	69.9	276	4	US-09-660-587-44
27	72	69.9	276	4	US-09-314-701-42
28	60	58.3	278	3	US-08-953-326-16
29	60	58.3	278	4	US-09-660-587-13
30	60	58.3	278	4	US-09-261-358A-13
31	60	58.3	278	4	US-09-201-458-9
32	60	58.3	278	4	US-09-314-701-10
33	60	58.3	280	4	US-09-660-587-11
34	60	58.3	280	4	US-09-261-358A-11
35	60	58.3	280	4	US-09-201-458-7
36	60	58.3	280	4	US-09-314-701-6
37	59.5	57.8	287	3	US-08-733-230-2
38	59.5	57.8	287	3	US-08-953-326-2
39	59	57.3	284	4	US-09-660-587-15
40	59	57.3	284	4	US-09-261-358A-15
41	59	57.3	284	4	US-09-201-458-11
42	55	53.4	133	3	US-08-953-326-20
43	55	53.4	133	4	US-09-660-587-7
44	55	53.4	133	4	US-09-261-358A-7
45	55	53.4	133	4	US-09-201-458-3
46	55	53.4	283	4	US-09-660-587-4
47	55	53.4	283	4	US-09-261-358A-4
48	44	42.7	308	4	US-09-584-568C-8
49	43.5	42.2	268	4	US-09-252-991A-21630
50	43	41.7	430	4	US-09-252-991A-26356
51	43	41.7	540	4	US-09-513-057C-33
52	42	40.8	436	4	US-09-252-991A-28754
53	42	40.8	447	4	US-09-252-991A-25916
54	41	39.8	135	1	US-08-446-600A-4
55	40	38.8	299	4	US-09-584-568C-6
56	40	38.8	513	4	US-09-252-991A-18177
57	40	38.8	1025	2	US-08-304-309-2
58	40	38.8	1025	3	US-08-991-942-2
59	40	38.8	1025	3	US-09-138-103-2
60	40	38.8	1025	5	PCT-US95-04567-4
61	39	37.9	1025	4	US-09-328-352-5300
62	39	37.9	1025	2	US-08-304-309-4
63	39	37.9	1025	3	US-08-991-942-4
64	39	37.9	1025	5	PCT-US95-04567-2
65	38	36.9	122	4	US-09-252-991A-16604
66	38	36.9	247	4	US-09-199-637A-363
67	38	36.9	247	4	US-09-252-991A-21412
68	38	36.9	280	4	US-09-660-587-42
69	38	36.9	280	4	US-09-314-701-48
70	38	36.9	283	4	US-09-660-587-10
71	38	36.9	283	4	US-09-261-358A-10
72	38	36.9	283	4	US-09-201-458-6
73	38	36.9	283	4	US-09-314-701-4
74	38	36.9	305	4	US-09-339-159B-18
75	38	36.9	371	3	US-08-586-165-3
76	38	36.9	372	3	US-08-586-165-5
77	38	36.9	421	4	US-09-328-352-8035
78	38	36.9	421	4	US-09-252-991A-17660
79	38	36.9	789	4	US-09-252-991A-27011
80	38	36.9	859	1	US-08-053-614-2
81	38	36.9	859	1	US-08-316-397B-2
82	38	36.9	859	2	US-09-034-306-2
83	38	36.9	859	2	US-09-259-437-2
84	38	36.9	859	5	PCT-US93-09782-2
85	38	36.9	956	4	US-09-134-078-63
86	38	36.9	1001	4	US-09-252-991A-26291
87	38	36.9	1181	1	US-08-053-614-4
88	38	36.9	1181	1	US-08-316-397B-4
89	38	36.9	1181	2	US-09-034-306-4
90	38	36.9	1181	3	US-09-259-437-4
91	38	36.9	1181	5	PCT-US93-09782-4
92	38	36.9	1233	3	US-09-328-352-7874
93	38	36.9	1255	3	US-08-947-823-3
94	38	36.9	1257	3	US-08-947-823-5
95	37.5	36.4	514	4	US-09-134-001C-4347
96	37.5	36.4	567	4	US-09-328-352-5912
97	37	35.9	31	1	US-08-190-802A-70
98	37	35.9	31	1	US-08-190-802A-129
99	37	35.9	31	1	US-08-190-802A-173
100	37	35.9	31	3	US-08-477-346-70

; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match 94.2%; Score 97; DB 3; Length 280;
Best Local Similarity 94.7%; Pred. No. 2.2e-08;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWDGSGTIS 19
||| |||||||||
DB 60 NNTTGVFGLKQDWDGSGTIS 78

RESULT 6

US-09-314-701-32
; Sequence 32, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikhhisa, Yasuko
; APPLICANT: Ohasi, No. 654451710
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-32

Query Match 93.2%; Score 96; DB 4; Length 288;
Best Local Similarity 94.4%; Pred. No. 3.4e-08;
Matches 17; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWDGSGTI 18
||| |||||||||
DB 60 NNTTGVFGLKQDWDGSGTI 77

RESULT 7

US-08-953-326-18
; Sequence 18, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17

; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match 85.4%; Score 88; DB 3; Length 276;
Best Local Similarity 84.2%; Pred. No. 6.9e-07;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWDGSGTIS 19
||| |||||||||
DB 59 NNTTGVFGLKQDWDGSAIS 77

RESULT 8

US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UF-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-4

Query Match 85.4%; Score 88; DB 3; Length 280;
Best Local Similarity 84.2%; Pred. No. 7e-07;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWDGSGTIS 19
||| |||||||||
DB 60 NNTTGVFGLKQDWDGSAIS 78

RESULT 9

US-08-953-326-4
; Sequence 4, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-4

Query Match 85.4%; Score 88; DB 3; Length 280;
Best Local Similarity 84.2%; Pred. No. 7e-07; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2;

QY 1 NTTGVFGLKQDWDGSTS 19
Db 60 NTTGVFGLKQDWDGSAIS 78
||| |||||:|||| ||

RESULT 10
US-09-660-587-9
; Sequence 9, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-660-587-9

Query Match 85.4%; Score 88; DB 4; Length 281;
Best Local Similarity 84.2%; Pred. No. 7e-07; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2;

QY 1 NTTGVFGLKQDWDGSTS 19
Db 59 NTTGVFGLKQDWDGSAIS 77
||| |||||:|||| ||

RESULT 11
US-09-261-358A-9
; Sequence 9, Application US/09261358A

; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 9
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-261-358A-9

Query Match 85.4%; Score 88; DB 4; Length 281;
Best Local Similarity 84.2%; Pred. No. 7e-07; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2;

QY 1 NTTGVFGLKQDWDGSTS 19
Db 59 NTTGVFGLKQDWDGSAIS 77
||| |||||:|||| ||

RESULT 12
US-09-201-458-5
; Sequence 5, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 5
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-201-458-5

Query Match 85.4%; Score 88; DB 4; Length 281;
Best Local Similarity 84.2%; Pred. No. 7e-07; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2;

QY 1 NTTGVFGLKQDWDGSTS 19
Db 59 NTTGVFGLKQDWDGSAIS 77
||| |||||:|||| ||

RESULT 13
US-09-314-701-2
; Sequence 2, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19

; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-2

Query Match 85.4%; Score 88; DB 4; Length 281;
Best Local Similarity 84.2%; Pred. No. 7e-07; 2; Indels 0; Gaps 0;
Matches 16; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQWDGSGTIS 19
||| ||||| ||||| ||
DB 59 NSTGVFGLKQWDGSAIS 77

RESULT 14
US-09-660-587-6
; Sequence 6, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-6 protein
US-09-660-587-6

Query Match 77.7%; Score 80; DB 4; Length 280;
Best Local Similarity 73.7%; Pred. No. 1.5e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQWDGSGTIS 19
||| ||||| ||||| ||
DB 59 NSTGVFGLKQWDGSGTIS 77

RESULT 15
US-09-261-358A-6
; Sequence 6, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28SA3 protein
US-09-261-358A-6

Query Match 77.7%; Score 80; DB 4; Length 280;
Best Local Similarity 73.7%; Pred. No. 1.5e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQWDGSGTIS 19
||| ||||| ||||| ||
DB 59 NSTGVFGLKQWDGSGTIS 77

RESULT 16
US-09-314-701-38
; Sequence 38, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohsai, No. 6544517lo
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-38

Query Match 77.7%; Score 80; DB 4; Length 280;
Best Local Similarity 73.7%; Pred. No. 1.5e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQWDGSGTIS 19
||| ||||| ||||| ||
DB 59 NSTGVFGLKQWDGSGTIS 77

RESULT 17
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nvika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 71.8%; Score 74; DB 3; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQWDGSGTIS 19

```
Db      60 NTVGVFGIEQDWDRCVIS 78
|||||:|||||
RESULT 18
US-09-660-587-12
; Sequence 12, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1D
US-09-660-587-12
Query Match      71.8%; Score 74; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      1 NNTTGVFGLKODWDGSGTIS 19
Db      60 NTVGVFGIEQDWDRCVIS 78
|||||:|||||

RESULT 19
US-09-261-358A-12
; Sequence 12, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR FILING DATE: 09/201,458
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1D
US-09-261-358A-12
Query Match      71.8%; Score 74; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      1 NNTTGVFGLKODWDGSGTIS 19
Db      60 NTVGVFGIEQDWDRCVIS 78
|||||:|||||

RESULT 20
US-09-201-458-8
; Sequence 8, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1D
US-09-201-458-8
Query Match      71.8%; Score 74; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      1 NNTTGVFGLKODWDGSGTIS 19
Db      60 NTVGVFGIEQDWDRCVIS 78
|||||:|||||

RESULT 21
US-09-314-701-8
; Sequence 8, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Kikihisa, Yasuko
; APPLICANT: Ogasai, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: chaffeensis
; FILE REFERENCE: 72727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1D
US-09-314-701-8
Query Match      71.8%; Score 74; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      1 NNTTGVFGLKODWDGSGTIS 19
Db      60 NTVGVFGIEQDWDRCVIS 78
|||||:|||||

RESULT 22
US-09-660-587-2
; Sequence 2, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR FILING DATE: 09/261,358
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1D
US-09-660-587-2
Query Match      71.8%; Score 74; DB 4; Length 286;
Best Local Similarity 68.4%; Pred. No. 0.00015;
Matches 13; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      1 NNTTGVFGLKODWDGSGTIS 19
Db      60 NTVGVFGIEQDWDRCVIS 78
|||||:|||||
```

; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein
US-09-660-587-2

Query Match 70.9%; Score 73; DB 4; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00022;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFLKQDWGSTI 18
:| ||||| ||||| |
Db 61 STGVFLKHDWDGSP1 77

RESULT 23
US-09-261-358A-2
; Sequence 2, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28-1 protein
US-09-261-358A-2

Query Match 70.9%; Score 73; DB 4; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00022;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFLKQDWGSTI 18
:| ||||| ||||| |
Db 61 STGVFLKHDWDGSP1 77

RESULT 24
US-09-201-458-2
; Sequence 2, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of a 30 kDa immunoreactive
US-09-201-458-2

Query Match 70.9%; Score 73; DB 4; Length 278;
Best Local Similarity 76.5%; Pred. No. 0.00022;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFLKQDWGSTI 18
:| ||||| ||||| |
Db 61 STGVFLKHDWDGSP1 77

RESULT 25
US-09-314-701-36
; Sequence 36, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-36

Query Match 70.9%; Score 73; DB 4; Length 307;
Best Local Similarity 76.5%; Pred. No. 0.00024;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 TTTGVFLKQDWGSTI 18
:| ||||| ||||| |
Db 90 STGVFLKHDWDGSP1 106

RESULT 26
US-09-660-587-44
; Sequence 44, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 44
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-3 protein
US-09-660-587-44

Query Match 69.9%; Score 72; DB 4; Length 276;
Best Local Similarity 66.7%; Pred. No. 0.00031;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NTTTGVFLKQDWGSTI 18
|||||:||||:| |
Db 60 NTTTGIFLKESWTGGII 77

RESULT 27
US-09-314-701-42

```
; Sequence 42, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. 6544517a
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: D6152CIP
; PRIOR FILING DATE: 1997-03-03
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-42

Query Match          69.9%; Score 72; DB 4; Length 278;
Best Local Similarity 66.7%; Pred. No. 0.00031;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGSGTI 18
DB 60 NNTTGVFGLKQDWGSGTI 77

RESULT 28
US-08-953-326-16
; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Ronan R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match          58.3%; Score 60; DB 3; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGSGTI 19
DB 60 NPTVALYGLKQDWEGISS 78

RESULT 29
US-09-660-587-13
; Sequence 13, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9
```

```
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 1999-03-03
; PRIOR FILING DATE: 1998-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-660-587-13

Query Match          58.3%; Score 60; DB 4; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGSGTI 19
DB 60 NPTVALYGLKQDWEGISS 78

RESULT 30
US-09-261-358A-13
; Sequence 13, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR FILING DATE: 09/201,458
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-261-358A-13

Query Match          58.3%; Score 60; DB 4; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NNTTGVFGLKQDWGSGTI 19
DB 60 NPTVALYGLKQDWEGISS 78

RESULT 31
US-09-201-458-9
; Sequence 9, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 9
```

; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-201-458-9

Query Match 58.3%; Score 60; DB 4; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTIS 19
| | :|||||: |
Db 60 NPTVALYGLKQDWDGSISS 78

RESULT 32
US-09-314-701-10
; Sequence 10, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; PRIOR FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-10

Query Match 58.3%; Score 60; DB 4; Length 278;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTIS 19
| | :|||||: |
Db 60 NPTVALYGLKQDWDGSISS 78

RESULT 33
US-09-660-587-11
; Sequence 11, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-660-587-11

Query Match 58.3%; Score 60; DB 4; Length 280;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTIS 19
| | :|||||: |
Db 60 NPTVALYGLKQDWDGVSAS 78

RESULT 34
US-09-261-358A-11
; Sequence 11, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 11
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-261-358A-11

Query Match 58.3%; Score 60; DB 4; Length 280;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTIS 19
| | :|||||: |
Db 60 NPTVALYGLKQDWDGVSAS 78

RESULT 35
US-09-201-458-7
; Sequence 7, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 7
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1C
US-09-201-458-7

Query Match 58.3%; Score 60; DB 4; Length 280;
Best Local Similarity 52.6%; Pred. No. 0.031;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 NTTGVFGLKQDWDGSTIS 19
| | :|||||: |
Db 60 NPTVALYGLKQDWDGVSAS 78

Search completed: October 6, 2003, 07:53:55
Job time : 11.0597 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 6, 2003, 07:43:07 : Search time 11.6418 seconds
(without alignments)
72.688 Million cell updates/sec

Title: US-09-765-739A-1
Perfect score: 110
Sequence: 1 KSTGVFGFKHWDGSPILK 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep: *
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	100.0	278	4	US-09-660-587-2
2	110	100.0	278	4	US-09-261-358A-2
3	110	100.0	278	4	US-09-201-458-2
4	110	100.0	307	4	US-09-314-701-36
5	79	71.8	280	3	US-08-953-326-17
6	79	71.8	280	4	US-09-660-587-6
7	79	71.8	280	4	US-09-261-358A-6
8	79	71.8	280	4	US-09-314-701-38
9	76	69.1	280	4	US-09-660-587-14
10	76	69.1	280	4	US-09-261-358A-14
11	76	69.1	280	4	US-09-201-458-10
12	76	69.1	280	4	US-09-314-701-12
13	72	65.5	276	3	US-08-953-326-18
14	72	65.5	280	3	US-08-733-230-4
15	72	65.5	280	3	US-08-953-326-4
16	72	65.5	281	4	US-09-660-587-9
17	72	65.5	281	4	US-09-261-358A-9
18	72	65.5	281	4	US-09-201-458-5
19	72	65.5	281	4	US-09-314-701-2
20	70	63.6	288	4	US-09-314-701-32
21	60	54.5	287	3	US-08-733-230-2
22	60	54.5	287	3	US-08-953-326-2
23	58	52.7	284	4	US-09-660-587-15
24	58	52.7	284	4	US-09-261-358A-15
25	58	52.7	284	4	US-09-201-458-11
26	58	52.7	286	3	US-08-953-326-15
27	58	52.7	286	4	US-09-660-587-12
28	58	52.7	286	4	US-09-261-358A-12
29	58	52.7	286	4	US-09-201-458-8
30	58	52.7	286	4	US-09-314-701-8
31	57	51.8	276	4	US-09-660-587-44
32	57	51.8	276	4	US-09-314-701-42
33	53	48.2	278	3	US-08-953-326-16
34	53	48.2	278	4	US-09-660-587-13
35	53	48.2	278	4	US-09-261-358A-13
36	53	48.2	278	4	US-09-201-458-9
37	53	48.2	278	4	US-09-314-701-10
38	52	47.3	280	4	US-09-660-587-11
39	52	47.3	280	4	US-09-261-358A-11
40	52	47.3	280	4	US-09-201-458-7
41	52	47.3	280	4	US-09-314-701-6
42	48	43.6	133	3	US-08-953-326-20
43	48	43.6	133	4	US-09-660-587-7
44	48	43.6	133	4	US-09-261-358A-7
45	48	43.6	133	4	US-09-201-458-3
46	48	43.6	283	4	US-09-660-587-4
47	48	43.6	283	4	US-09-261-358A-4
48	46	41.8	859	1	US-08-053-614-2
49	46	41.8	859	1	US-08-316-397B-2
50	46	41.8	859	2	US-09-034-306-2
51	46	41.8	859	3	US-09-259-437-2
52	46	41.8	859	5	PCT-US93-09782-2
53	46	41.8	1181	1	US-08-053-614-4
54	46	41.8	1181	1	US-08-316-397B-4
55	46	41.8	1181	2	US-09-034-306-4
56	46	41.8	1181	3	US-09-259-437-4
57	46	41.8	1181	5	PCT-US93-09782-4
58	45	40.9	562	4	US-09-369-247-144
59	45	40.9	570	1	US-08-403-866-6
60	44	40.0	327	4	US-09-252-991A-25901
61	43	39.1	136	3	US-09-053-197A-37
62	43	39.1	136	4	US-09-085-761A-37
63	43	39.1	137	4	US-09-134-001C-5560
64	43	39.1	265	4	US-09-134-001C-3336
65	43	39.1	502	4	US-09-252-991A-23390
66	43	39.1	513	4	US-09-252-991A-18177
67	42	38.2	305	4	US-09-252-991A-25874
68	42	38.2	428	4	US-09-054-680-4
69	42	38.2	449	4	US-09-252-991A-17901
70	42	38.2	592	2	US-08-366-490-2
71	42	38.2	592	2	US-08-366-490-6
72	42	38.2	592	3	US-08-860-483A-2
73	42	38.2	592	3	US-08-860-483A-6
74	42	38.2	596	3	US-08-860-483A-12
75	42	38.2	596	3	US-08-860-483A-13
76	42	38.2	632	2	US-08-366-490-8
77	41	37.3	155	4	US-09-252-991A-25374
78	41	37.3	437	4	US-09-252-991A-29909
79	41	37.3	890	3	US-09-342-648-10
80	40	36.4	396	3	US-08-860-656B-2
81	40	36.4	396	3	US-08-860-656B-3
82	40	36.4	396	3	US-08-860-656B-4
83	40	36.4	529	1	US-08-152-019A-40
84	40	36.4	529	3	US-08-650-599A-3
85	40	36.4	529	4	US-09-490-517-3
86	40	36.4	557	4	US-09-252-991A-25674
87	40	36.4	569	2	US-08-467-822-27
88	40	36.4	569	3	US-08-432-697-27
89	40	36.4	569	3	US-08-466-248-27
90	40	36.4	603	3	US-08-482-677-8
91	40	36.4	604	2	US-08-635-137-2
92	40	36.4	604	3	US-09-136-981-2
93	40	36.4	605	1	US-08-152-019A-26
94	40	36.4	605	3	US-08-482-677-4
95	40	36.4	605	3	US-08-650-599A-1
96	40	36.4	605	4	US-09-490-517-1
97	40	36.4	606	3	US-08-665-259-22
98	40	36.4	606	3	US-08-762-500-22
99	40	36.4	606	3	US-09-306-902A-9
100	40	36.4	1001	4	US-09-252-991A-26291

ALIGNMENTS

```

RESULT 1
US-09-660-587-2
; Sequence 2, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152C1P2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-7 protein
US-09-660-587-2

```

Query Match 100.0%; Score 110; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 KSTVGFLKHDWDGSPILK 20
        |||||
Db      60 KSTVGFLKHDWDGSPILK 79
```

```

RESULT 2
US-09-261-358A-2
; Sequence 2, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152C1P
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33

```

```

; SEQ ID NO 2
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:

```

OTHER INFORMATION: amino acid sequence of Eca28-1 protein
US-09-261-358A-2

```
Query Match      100.0%; Score 110; DB 4; Length 278;
Best Local Similarity 100.0%; Pred. No. 7e-11;
Matches 20; Conservative 0; Mismatches 0; Indels
```

QY 1 KSTVGVEGLKHDWDGSPILK 20
Db 60 KSTVGVEGLKHDWDGSPILK 79

RESULT 3
US-09-201-458-2

```

: Sequence 2, Application US/09201458A
: Patent No. 6,559,942
: GENERAL INFORMATION:
: APPLICANT: Walkerc David H.
: APPLICANT: McBride Jere W.
: APPLICANT: Yu Yue-Jie
: TITLE OF INVENTION: 28-kDa Immureactive Protein Gene of Ehrlichia
: TITLE OF INVENTION: canis and Uses Thereof
: FILE REFERENCE: D6152
: CURRENT APPLICATION NUMBER: US/09/201.458A
: CURRENT FILING DATE: 1998-11-30
: NUMBER OF SEQ ID NOS: 21
: SEQ ID NO 2
: LENGTH: 278
: TYPE: PRT
: ORGANISM: Ehrlichia canis
: FEATURE:
: OTHER INFORMATION: amino acid sequence of a 30 kDa immunoreactive
: OTHER INFORMATION: protein of Ehrlichia canis
: US-09-201-458-2

```

Query Match	100.0%	Score 110;	DB 4;	Length 278;
Best Local Similarity	100.0%;	Pred. No. 7e-11;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	KSTVG	FGLK	KHD	WDG	SPILK	20
D _b	60	KSTVG <td>FGLK <td>KHD <td>WDG <td>SPILK <td>79</td> </td></td></td></td>	FGLK <td>KHD <td>WDG <td>SPILK <td>79</td> </td></td></td>	KHD <td>WDG <td>SPILK <td>79</td> </td></td>	WDG <td>SPILK <td>79</td> </td>	SPILK <td>79</td>	79

RESULT 4

RESOL 4
US-09-314-701-36
; Sequence 36, Application US/09314701

```

: Patent No. 6544517
:
: GENERAL INFORMATION:
:
: APPLICANT: Rkihisa, Yasuko
: APPLICANT: Onasi, No. 6544517to
: TITLE OF INVENTION: Outer Membr
: TITLE OF INVENTION: Chaffensis
: FILE REFERENCE: 22727/04021
: CURRENT APPLICATION NUMBER: US/
: CURRENT FILING DATE: 1999-05-19
: NUMBER OF SEQ ID NOS: 66
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 36
: LENGTH: 307

```

TYPE: FRI
ORGANISM: Ehrlichia canis
US-09-314-701-36

```
Query Match      100.0%; Score 110; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 7.8e-11;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	20
Qy	KSTVGVFGLKHDWDGSPILK	20
Db	KSTVGVFGLKHDWDGSPILK	108

RESULT 5

US-08-953-326-17
; Sequence 17, Application US/08953326
; Patent No. 6251872

; GENERAL INFORMATION:
 ; APPLICANT: Garbet, Anthony F.
 ; APPLICANT: Ganta, Roman R.
 ; APPLICANT: McGuire, Travis C.
 ; APPLICANT: Burridge, Michael J.
 ; APPLICANT: Nyika, Aceme
 ; APPLICANT: Rurangirwa, Fred R.

; APPLICANT: MARSH, SUMMIT M.
 ; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
 ; TITLE OF INVENTION: Animals and Humans

; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-17

Query Match 71.8%; Score 79; DB 3; Length 280;
Best Local Similarity 73.7%; Pred. No. 1.3e-05;
Matches 14; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPILK 20
Db 61 TTIGVFGLKODWDGWSISK 79
:1:||||| ||||| |

RESULT 6
US-09-660-587-6
; Sequence 6, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-6 protein
US-09-660-587-6

Query Match 71.8%; Score 79; DB 4; Length 280;
Best Local Similarity 82.4%; Pred. No. 1.3e-05;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPIL 18
Db 60 STGVFGLKHDWNGGTI 76
||||| ||||| |

RESULT 7
US-09-261-358A-6
; Sequence 6, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33

; SEQ ID NO 6
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28SA3 protein
US-09-261-358A-6

Query Match 71.8%; Score 79; DB 4; Length 280;
Best Local Similarity 82.4%; Pred. No. 1.3e-05;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPIL 18
Db 60 STGVFGLKHDWNGGTI 76
||||| ||||| |

RESULT 8
US-09-314-701-38
; Sequence 38, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-38

Query Match 71.8%; Score 79; DB 4; Length 280;
Best Local Similarity 82.4%; Pred. No. 1.3e-05;
Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 STGVFGLKHDWDGSPIL 18
Db 60 STGVFGLKHDWNGGTI 76
||||| ||||| |

RESULT 9
US-09-660-587-14
; Sequence 14, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-660-587-14

Query Match 69.1%; Score 76; DB 4; Length 280;
Best Local Similarity 73.7%; Pred. No. 4.2e-05;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
QY      2 STVGVLKHDWDGSPILK 20
DB      61 TTTGVFLKQDWDGSTISK 79

RESULT 10
US-09-261-358A-14
; Sequence 14, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 14
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-261-358A-14

Query Match      69.1%; Score 76; DB 4; Length 280;
Best Local Similarity 73.7%; Pred. No. 4.2e-05;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      2 STVGVLKHDWDGSPILK 20
DB      61 TTTGVFLKQDWDGSTISK 79

RESULT 11
US-09-201-458-10
; Sequence 10, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 10
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1F
US-09-201-458-10

Query Match      69.1%; Score 76; DB 4; Length 280;
Best Local Similarity 73.7%; Pred. No. 4.2e-05;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      2 STVGVLKHDWDGSPILK 20
DB      61 TTTGVFLKQDWDGSTISK 79

RESULT 12
US-09-314-701-12
; Sequence 12, Application US/09314701
```

```
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Chasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-12

Query Match      69.1%; Score 76; DB 4; Length 280;
Best Local Similarity 73.7%; Pred. No. 4.2e-05;
Matches 14; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      2 STVGVLKHDWDGSPILK 20
DB      61 TTTGVFLKQDWDGSTISK 79

RESULT 13
US-08-953-326-18
; Sequence 18, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman W.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-18

Query Match      65.5%; Score 72; DB 3; Length 276;
Best Local Similarity 76.5%; Pred. No. 0.0002;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 STVGVLKHDWDGSPILK 18
DB      60 TTTGVFLKQDWDGSAI 76

RESULT 14
US-08-733-230-4
; Sequence 4, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
```

APPLICANT: Nyika, Aceme
APPLICANT: Rurangirwa, Fred R.
APPLICANT: Mahan, Suman M.
TITLE OF INVENTION: Nucleic Acid Vaccines Against
TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,230
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Whitlock, Ted W.
REGISTRATION NUMBER: 36,965
REFERENCE/DOCKET NUMBER: UF-167
TELECOMMUNICATION INFORMATION:
TELEPHONE: 352-375-8100
TELEFAX: 352-372-5800
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-733-230-4

Query Match 65.5%; Score 72; DB 3; Length 280;
Best Local Similarity 76.5%; Pred. No. 0.0002;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
:|||||||:|||||
Db 61 TTGVFGLKQNDGSAI 77

RESULT 15
US-08-953-326-4
Sequence 4, Application US/08953326
Patent No. 6251872
GENERAL INFORMATION:
APPLICANT: Barbet, Anthony F.
APPLICANT: Ganta, Roman R.
APPLICANT: McGuire, Travis C.
APPLICANT: Burridge, Michael J.
APPLICANT: Nyika, Aceme
APPLICANT: Rurangirwa, Fred R.
TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
TITLE OF INVENTION: Animals and Humans
FILE REFERENCE: UF-167C1
CURRENT APPLICATION NUMBER: US/08/953,326
CURRENT FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/953,326
EARLIER FILING DATE: 1997-10-17
EARLIER APPLICATION NUMBER: 08/733,230
EARLIER FILING DATE: 1996-10-17
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 280
TYPE: PRT
ORGANISM: Ehrlichia chaffeensis

US-08-953-326-4

Query Match 65.5%; Score 72; DB 3; Length 280;
Best Local Similarity 76.5%; Pred. No. 0.0002;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
:|||||||:|||||
Db 61 TTGVFGLKQNDGSAI 77

RESULT 16
US-09-660-587-9
Sequence 9, Application US/09660587
Patent No. 6392023
GENERAL INFORMATION:
APPLICANT: Walker, David H.
APPLICANT: McBride, Jere W.
APPLICANT: Yu, Xue-Jie
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
FILE REFERENCE: D6152CIP2
CURRENT APPLICATION NUMBER: US/09/660,587
CURRENT FILING DATE: 2000-09-12
PRIOR APPLICATION NUMBER: 09/261,358
PRIOR FILING DATE: 1999-03-03
NUMBER OF SEQ ID NOS: 46
SEQ ID NO 9
LENGTH: 281
TYPE: PRT
ORGANISM: Ehrlichia chaffeensis
FEATURE:
OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-660-587-9

Query Match 65.5%; Score 72; DB 4; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.0002;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18
:|||||||:|||||
Db 60 TTGVFGLKQNDGSAI 76

RESULT 17
US-09-261-358A-9
Sequence 9, Application US/09261358A
Patent No. 6403780
GENERAL INFORMATION:
APPLICANT: Walker, David H.
APPLICANT: McBride, Jere W.
APPLICANT: Yu, Xue-Jie
TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
FILE REFERENCE: D6152CIP
CURRENT APPLICATION NUMBER: US/09/261,358A
CURRENT FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: 09/201,458
PRIOR FILING DATE: 1998-11-30
NUMBER OF SEQ ID NOS: 33
SEQ ID NO 9
LENGTH: 281
TYPE: PRT
ORGANISM: Ehrlichia chaffeensis
FEATURE:
OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-261-358A-9

Query Match 65.5%; Score 72; DB 4; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.0002;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 STVGVEGLKHDWDGSP1 18

```
Db      60 TTGVFGLKQNDGSAI 76
:|||||:|||||
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-32

Query Match      63.6%; Score 70; DB 4; Length 288;
Best Local Similarity 70.6%; Pred. No. 0.00046;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      2 STGVFGLKHDWDGSP1 18
:|||||:|||||
Db      61 TTGVFGLKQNDGATI 77
:|||||:|||||

RESULT 21
US-08-733-230-2
; Sequence 2, Application US/08733230
; Patent No. 6025338
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman Reddy
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against
; TITLE OF INVENTION: Rickettsial Diseases and Methods of Use
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08733,230
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitlock, Ted W.
; REGISTRATION NUMBER: 36,965
; REFERENCE/DOCKET NUMBER: UP-167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-733-230-2

Query Match      54.5%; Score 60; DB 3; Length 287;
Best Local Similarity 73.3%; Pred. No. 0.023;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 KSTGVFGLKHDWDG 15
:|||||:|||||
Db      61 KNTQVFLKQNDWG 75
:|||||:|||||

; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis P28
US-09-201-458-5

Query Match      65.5%; Score 72; DB 4; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.0002;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 STGVFGLKHDWDGSP1 18
:|||||:|||||
Db      60 TTGVFGLKQNDGSAI 76
:|||||:|||||

RESULT 19
US-09-314-701-2
; Sequence 2, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517a
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-2

Query Match      65.5%; Score 72; DB 4; Length 281;
Best Local Similarity 76.5%; Pred. No. 0.0002;
Matches 13; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 STGVFGLKHDWDGSP1 18
:|||||:|||||
Db      60 TTGVFGLKQNDGSAI 76
:|||||:|||||

RESULT 20
US-09-314-701-32
; Sequence 32, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517a
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeensis
```

RESULT 22
US-08-953-326-2
; Sequence 2, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nyika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
US-08-953-326-2

Query Match 54.5%; Score 60; DB 3; Length 287;
Best Local Similarity 73.3%; Pred. No. 0.023;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSTVGFGGLKHDWDG 15
; : | ||||| ||||
Db 61 KNTQTVEGLKKDWDG 75

RESULT 23
US-09-660-587-15
; Sequence 15, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-660-587-15

Query Match 52.7%; Score 58; DB 4; Length 284;
Best Local Similarity 66.7%; Pred. No. 0.049;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KSTVGFGGLKHDWDG 15
; : | ||||| ||||
Db 61 RDTKAVFGLKKDWDG 75

RESULT 24
US-09-261-358A-15
; Sequence 15, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 15
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-261-358A-15

Query Match 52.7%; Score 58; DB 4; Length 284;
Best Local Similarity 66.7%; Pred. No. 0.049;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KSTVGFGGLKHDWDG 15
; : | ||||| ||||
Db 61 RDTKAVFGLKKDWDG 75

RESULT 25
US-09-201-458-11
; Sequence 11, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; TITLE OF INVENTION: canis and Uses Thereof
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 11
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Cowdria ruminantium
; FEATURE:
; OTHER INFORMATION: amino acid sequence of C. ruminantium MAP-1
US-09-201-458-11

Query Match 52.7%; Score 58; DB 4; Length 284;
Best Local Similarity 66.7%; Pred. No. 0.049;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KSTVGFGGLKHDWDG 15
; : | ||||| ||||
Db 61 RDTKAVFGLKKDWDG 75

RESULT 26
US-08-953-326-15
; Sequence 15, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.

; APPLICANT: Nvika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-15

Query Match 52.7%; Score 58; DB 3; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.05;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 STGVFGLKHDWD 14
Db 61 TTGVFGIEQDWD 73

RESULT 27
US-09-660-587-12
; Sequence 12, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-660-587-12

Query Match 52.7%; Score 58; DB 4; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.05;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 STGVFGLKHDWD 14
Db 61 TTGVFGIEQDWD 73

RESULT 28
US-09-261-358A-12
; Sequence 12, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP

; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 12
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-261-358A-12

Query Match 52.7%; Score 58; DB 4; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.05;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 STGVFGLKHDWD 14
Db 61 TTGVFGIEQDWD 73

RESULT 29
US-09-201-458-8
; Sequence 8, Application US/09201458A
; Patent No. 6458942
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: 28-kDa Immunoreactive Protein Gene of Ehrlichia
; FILE REFERENCE: D6152
; CURRENT APPLICATION NUMBER: US/09/201,458A
; CURRENT FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 21
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-ID
US-09-201-458-8

Query Match 52.7%; Score 58; DB 4; Length 286;
Best Local Similarity 69.2%; Pred. No. 0.05;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 STGVFGLKHDWD 14
Db 61 TTGVFGIEQDWD 73

RESULT 30
US-09-314-701-8
; Sequence 8, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohashi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-09-314-701-8

Query Match 52.7%; Score 58; DB 4; Length 286;
Best Local Similarity 59.2%; Pred. No. 0.05;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDND 14
Db 61 TTVGVFGIEQDWD 73

RESULT 31

US-09-660-587-44
; Sequence 44, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 44
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. canis p28-3 protein
US-09-660-587-44

Query Match 51.8%; Score 57; DB 4; Length 276;
Best Local Similarity 55.6%; Pred. No. 0.071;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSPIL 19
Db 61 TTTGIFGLKESWTGGIIL 78

RESULT 32

US-09-314-701-42
; Sequence 42, Application US/09314701
; Patent No. 6544517
; GENERAL INFORMATION:
; APPLICANT: Rikihisa, Yasuko
; APPLICANT: Ohasi, No. 6544517io
; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
; TITLE OF INVENTION: Chaffeeensis
; FILE REFERENCE: 22727/04021
; CURRENT APPLICATION NUMBER: US/09/314,701
; CURRENT FILING DATE: 1999-05-19
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 276
; TYPE: PRT
; ORGANISM: Ehrlichia canis
US-09-314-701-42

Query Match 51.8%; Score 57; DB 4; Length 276;
Best Local Similarity 55.6%; Pred. No. 0.071;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 STVGVFGLKHDWDGSPIL 19
Db 61 TTTGIFGLKESWTGGIIL 78

RESULT 33

US-08-953-326-16

; Sequence 16, Application US/08953326
; Patent No. 6251872
; GENERAL INFORMATION:
; APPLICANT: Barbet, Anthony F.
; APPLICANT: Ganta, Roman R.
; APPLICANT: McGuire, Travis C.
; APPLICANT: Burridge, Michael J.
; APPLICANT: Nvika, Aceme
; APPLICANT: Rurangirwa, Fred R.
; APPLICANT: Mahan, Suman M.
; TITLE OF INVENTION: Nucleic Acid Vaccines Against Rickettsial Diseases of
; TITLE OF INVENTION: Animals and Humans
; FILE REFERENCE: UF-167C1
; CURRENT APPLICATION NUMBER: US/08/953,326
; CURRENT FILING DATE: 1997-10-17
; EARLIER FILING DATE: 1997-10-17
; EARLIER APPLICATION NUMBER: 08/733,230
; EARLIER FILING DATE: 1996-10-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
US-08-953-326-16

Query Match 48.2%; Score 53; DB 3; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.34;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGLKHDWDG 15
Db 62 TVALYGLKQDWDG 74

RESULT 34

US-09-660-587-13
; Sequence 13, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; CURRENT FILING DATE: 2000-09-12
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-660-587-13

Query Match 48.2%; Score 53; DB 4; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.34;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 TVGVFGLKHDWDG 15
Db 62 TVALYGLKQDWDG 74

RESULT 35

US-09-261-358A-13
; Sequence 13, Application US/09261358A
; Patent No. 6403780

; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CJP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; CURRENT FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 13
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of E. chaffeensis OMP-1E
US-09-261-358A-13

Query Match 48.2%; Score 53; DB 4; Length 278;
Best Local Similarity 61.5%; Pred. No. 0.34;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 3 TVGVFGLKHDWDG 15
II :III II:
Db 62 TVALYGLKQDWEG 74

Search completed: October 6, 2003, 07:53:48
Job time : 14.6418 secs